



# Estimation of the parameters of a Markov-modulated loss process in insurance

Armelle Guillou, Stéphane Loisel, Gilles Stupfler

## ► To cite this version:

Armelle Guillou, Stéphane Loisel, Gilles Stupfler. Estimation of the parameters of a Markov-modulated loss process in insurance. Insurance: Mathematics and Economics, 2013, 53, pp.388-404. 10.1016/j.insmatheco.2013.07.003 . hal-00589696

**HAL Id: hal-00589696**

**<https://hal.science/hal-00589696>**

Submitted on 30 Apr 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Estimation of the parameters of a Markov-modulated loss process in insurance

Armelle Guillou<sup>(1)</sup>, Stéphane Loisel<sup>(2)</sup> & Gilles Stupfler<sup>(1)</sup>

<sup>(1)</sup> Université de Strasbourg & CNRS, IRMA, UMR 7501, 7 rue René Descartes,  
67084 Strasbourg cedex, France

<sup>(2)</sup> Université de Lyon, Université Lyon 1, Institut de Science Financière et d'Assurances, 50  
avenue Tony Garnier, 69007 Lyon, France

**Abstract.** We present a new model of loss processes in insurance. The process is a couple  $(N, L)$  where  $N$  is a univariate Markov-modulated Poisson process (MMPP) and  $L$  is a multivariate loss process whose behaviour is driven by  $N$ . We prove the strong consistency of the maximum likelihood estimator of the parameters of this model, and present an EM algorithm to compute it in practice. The method is illustrated with simulations and real sets of insurance data.

**Keywords:** Markov-modulated Poisson process, maximum likelihood estimator, strong consistency, EM algorithm.

## 1 Introduction

A Markov-modulated Poisson process (MMPP) is a doubly stochastic Poisson process whose intensity is driven by a non-observable continuous-time Markov chain with finite state space. A comprehensive survey of the properties of MMPPs is given in [15]. Such processes are used to model communication networks (see [18, 21]), environmental phenomena as in [13], and the surplus of an insurance company as in [1]. It has then been crucial to develop methods to estimate the parameters of such processes. From a theoretical point of view, the strong consistency of the maximum likelihood estimator (MLE) for an MMPP is shown by Rydén in [32]; his proof is strongly influenced by [23], in which consistency for the MLE for general hidden Markov models (HMMs) is established. The properties of the MLE in this context have been extensively studied since Baum and Petrie [3]: in addition to consistency in [23], asymptotic normality was proved in [5]. Now, from a practical point of view, the MLE of an MMPP can be computed via several methods, see [14, 19, 28, 30, 35]. In the

latter paper, an EM algorithm is presented, and it has become the standard tool to estimate the MLE. For other references on EM algorithms, we refer the reader to Baum *et al.* [4], who first presented such algorithm for HMMs; recent surveys on EM algorithms include the monograph by McLachlan and Krishnan [27]. Other possible approaches include matching moments and covariance functions, see [17, 31], or maximizing a split-time likelihood, as introduced by Rydén in [33, 34], further studied by Vandekerckhove [36] in the context of hidden mixtures of Markov processes. In [25], Loisel suggested that correlation between lines of business of an insurance company could be caused by common shocks and modulation by a common Markovian environment process. Our goal is to extend the MLE approach to estimate the parameters of a process  $(N, L)$  where  $N$  is a univariate MMPP and  $L$  is a (possibly multivariate) loss process whose behavior is driven by  $N$ , in order to estimate the parameters of such a process in two real sets of insurance data. We also carry out a simulation study of loss processes for 2 and 3 lines of business modulated by a common environment process. Our results confirm that the method works quite well as long as the observation period contains enough changes of the Markovian environment process.

## 2 Model, assumptions and notation

We consider an MMPP  $(J, N)$ , where  $J$  is an irreducible continuous-time Markov process with generator  $L$  on the state space  $\{1, \dots, r\}$ , where  $r \in \mathbb{N} \setminus \{0\}$ , and  $N$  is a univariate counting process such that, when  $J$  is in state  $i$ ,  $N$  is a Poisson process with intensity  $\lambda_i$ . We further consider a loss process  $S = (S_1, \dots, S_n)$  (namely, the  $S_k$  are piecewise constant processes with nonnegative increments) whose behavior is driven by  $N$  in the following sense: assume that the  $S_k$  can only jump when  $N$  does, and that if  $N$  jumps at time  $t$  and if  $J$  is in state  $i$ , then a simultaneous jump of the processes  $S_{k_1}, \dots, S_{k_p}$  at time  $t$  occurs with probability  $p_i(e)$  where  $e = \{k_1, \dots, k_p\}$  is a subset of  $\{1, \dots, n\}$ . We then assume that the random variables  $E_s$ , such that the  $S_k$  with  $k \in E_s$  jumped (and only these) at the time of the  $s$ -th jump of  $N$ , are independent given the process  $(J, N)$ . Finally, assume that the value of the jump  $X_s$  has distribution  $\mathbb{P}_{\theta(i, e)}$ , where  $(\mathbb{P}_{\theta})_{\theta \in \Theta}$  is a parametric statistical model, that is

$$\mathbb{P}(X_s = x \mid J(\tau_s) = i, E_s = e) = \mathbb{P}_{\theta(i, e)}(\forall m, m \in e \Rightarrow X_m = x_m)$$

where  $\tau_s$  is the time of the  $s$ -th jump of  $N$ , with clearly  $x_m = 0$  if  $m \notin e$ . Note that this model can be seen as a common shock model as in [24]: it is assumed that given the process  $(J, N)$  and the sequence  $(E_s)$ , the  $X_s$  are independent random variables.

The context of our work is the following: let us assume that the process  $S$  has been observed until time  $T$ , so that the available data is:

1. The number  $r$  of states of  $J$ ;
2. The full knowledge of the processes  $N$  and  $S$  between time 0 and time  $T$ , both assumed to be times when  $N$  jumps.

The goal is to estimate the unknown parameters of the model, namely:

1. The elements  $\ell_{ij}$  of the transition intensity matrix  $L$  of  $J$ ;
2. The jump intensities  $\lambda_i$  of  $N$ ;
3. The probabilities  $p_i(e)$ , where  $e$  is a subset of  $\{1, \dots, n\}$ ;
4. The parameters  $\theta(i, e)$ .

Remark that the process  $J$  is not observed, which induces technical difficulties. For the sake of shortness, we let  $\Phi$  be the global parameter of the model. The distribution of the process with parameter  $\Phi$  is then denoted by  $\mathbb{P}_\Phi$ .

### 3 Asymptotic properties of the maximum likelihood estimator

Our aim is to estimate the parameters with a maximum likelihood estimator (MLE). Let then  $Y_i = \tau_i - \tau_{i-1}$  be the amount of time between the  $(i-1)$ -th and the  $i$ -th shock, and  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$ .

The available data is:

1. The values  $0 < t_1 < \dots < t_k = T$  of the  $\tau_i$ , *i.e.* the times when  $N$  jumps (equivalently, the inter-event times  $y_1, \dots, y_k$ , where  $y_j = t_j - t_{j-1}$ ,  $t_0 = 0$ );
2.  $e_1, \dots, e_k$  the successive values of the  $E_k$ ;
3.  $x_1, \dots, x_k$  the successive values of the jumps of  $S$ .

Let now

$$\begin{aligned} f_{ij}(t, \Phi) dt &:= \mathbb{P}_\Phi(T_1 \in dt, J(t) = j \mid J(0) = i) \\ \overline{F}_{ij}(t, \Phi) &:= \mathbb{P}_\Phi(T_1 > t, J(t) = j \mid J(0) = i). \end{aligned}$$

Therefore (see [28]), we have

$$f(t, \Phi) = \exp(t(L(\Phi) - \Lambda(\Phi)))\Lambda(\Phi), \quad \overline{F}(t, \Phi) = \exp(t(L(\Phi) - \Lambda(\Phi))).$$

Let then

$$\begin{aligned} p(e, \Phi) &= \text{diag}((p_i(e, \Phi))_{1 \leq i \leq r}), \\ \mathbb{P}_{\theta(\cdot, e, \Phi)}(X = x) &= \text{diag}((\mathbb{P}_{\theta(i, e, \Phi)}(X = x))_{1 \leq i \leq r}), \end{aligned}$$

and in matrix notation

$$\begin{aligned} \forall e \in \{1, \dots, n\}, e \neq \emptyset, \quad g(t, e, x, \Phi) &= f(t, \Phi) \cdot p(e, \Phi) \cdot \mathbb{P}_{\theta(\cdot, e, \Phi)}(X = x) \\ g(t, \emptyset, x, \Phi) &= f(t, \Phi) \cdot p(\emptyset, \Phi) \cdot \mathbb{1}_{\{x=0\}}. \end{aligned}$$

With these notations, the  $(i, j)$ -th element of the matrix  $g(t, e, x, \Phi)$  is

$$\begin{aligned} \forall e \in \{1, \dots, n\}, e \neq \emptyset, \quad g_{ij}(t, e, x, \Phi) &= f_{ij}(t, \Phi) p_j(e, \Phi) \mathbb{P}_{\theta(j, e, \Phi)}(X = x) \\ g_{ij}(t, \emptyset, x, \Phi) &= f_{ij}(t, \Phi) p_j(\emptyset, \Phi) \mathbb{1}_{\{x=0\}}. \end{aligned}$$

It is now sufficient to specify the starting distribution of  $J$  to compute the likelihood of the observations. Denote by  $P(\Phi)$  the transition matrix of the discrete-time Markov chain  $(J_i = J(\tau_i))$ : integrating  $f$ , one gets

$$P(\Phi) = (\Lambda(\Phi) - L(\Phi))^{-1} \Lambda(\Phi).$$

According to [32],  $P(\Phi)$  has a unique stationary distribution  $\pi(\Phi)$  and we have, if  $a(\Phi)$  is the only stationary distribution of the continuous-time process  $(J(t))_{t \geq 0}$  and  $\eta$  is the column vector of size  $r$  with all entries equal to 1,

$$\pi(\Phi) = \frac{1}{a(\Phi) \Lambda(\Phi) \eta} a(\Phi) \Lambda(\Phi).$$

We assume that the starting distribution of  $J$  is  $\pi(\Phi)$ ; the process  $((J_i, Y_i, E_i, X_i))_i$  is then  $\mathbb{P}_\Phi$ -stationary, because the bivariate process  $((J_i, Y_i))_i$  is a Markov renewal process (see e.g. [12, p. 313]). Thus, the likelihood of the observed data under the distribution  $\mathbb{P}_\Phi$  is

$$\mathcal{L}((y_i, e_i, x_i)_{1 \leq i \leq k}, \Phi) = \pi(\Phi) \left( \prod_{i=1}^k g(y_i, e_i, x_i, \Phi) \right) \eta.$$

Assuming now that we know the states  $j_0, j_1, \dots, j_k$  of the (hidden) Markov process  $J$  at the times when  $N$  jumps, the complete likelihood of the data is

$$\mathcal{L}((j_i)_{0 \leq i \leq k}, (y_i, e_i, x_i)_{1 \leq i \leq k}, \Phi) = \pi_{j_0}(\Phi) \left( \prod_{i=1}^k g_{j_{i-1}, j_i}(y_i, e_i, x_i, \Phi) \right).$$

To give a result on the strong consistency of the MLE, we first need some notations: for an arbitrary parameter  $\Phi$ , denote by  $F_\Phi$  the set of all parameters  $\Phi'$  such that for all  $e$

$$(\forall j \quad \lambda_j(\Phi) p_j(e, \Phi) = 0) \Leftrightarrow (\forall j \quad \lambda_j(\Phi') p_j(e, \Phi') = 0).$$

$F_\Phi$  can be thought of as the set of the elements  $\Phi'$  such that a simultaneous jump of the processes  $S_{k_1}, \dots, S_{k_q}$  is a.s. impossible under the law  $\mathbb{P}_\Phi$  if and only if it is a.s. impossible

under the law  $\mathbb{P}_{\Phi'}$ . Write further  $\Phi \sim \Phi'$  whenever  $((Y_i, E_i, X_i))_i$  has the same law under  $\mathbb{P}_{\Phi}$  and under  $\mathbb{P}_{\Phi'}$ .

We finally write down the hypotheses we need to state our main result:

- (A<sub>1</sub>) For all  $e \neq \emptyset$ , the distributions  $\mathbb{P}_{\theta(\cdot, e)}$  have the same support, with no atom at 0.
- (A<sub>2</sub>) For all  $e \neq \emptyset$  and all  $\Phi, \Phi'$ , there exists a neighborhood  $G$  of  $\Phi'$  such that for every subset  $G_{\Phi'}$  of  $G$  and all  $i, j \in \{1, \dots, r\}$ ,

$$\int \left| \ln \sup_{\varphi \in G_{\Phi'}} \mathbb{P}_{\theta(i, e, \varphi)}(m \in e \Rightarrow X_m = x_m) \right| \mathbb{P}_{\theta(j, e, \Phi)}(m \in e \Rightarrow X_m = x_m) dx < \infty.$$

- (A<sub>3</sub>) For all  $e \neq \emptyset$ , all  $i \in \{1, \dots, r\}$  and all  $x$ ,  $\varphi \mapsto \mathbb{P}_{\theta(i, e, \varphi)}(m \in e \Rightarrow X_m = x_m)$  is a continuous function.

This allows us to state our main result:

**Theorem 1.** *Assume that (A<sub>1</sub> – A<sub>3</sub>) hold. Let  $\Phi_0$  be the true value of the parameter, and let  $C$  be a compact set of  $F_{\Phi_0}$  such that  $\Phi_0 \in C$ . Let  $\hat{\Phi}_p$  be the MLE for  $\Phi_0$  on  $C$ , computed with  $p$  observations. Then if  $O \subset C$  is an open set in  $F_{\Phi_0}$  containing the equivalence class of  $\Phi_0$  modulo  $\sim$ , one has  $\hat{\Phi}_p \in C$  a.s. for  $p$  large enough.*

**Proof of Theorem 1.** We closely follow the proof of Theorem 1 in [32]: pick  $\Phi$  and  $\Phi' \in F_{\Phi_0}$  such that  $\Phi' \approx \Phi$ . Lemma 8 implies that there exists  $\varepsilon > 0$  such that  $H(\Phi, \Phi') < H(\Phi, \Phi) - 2\varepsilon$ . Now, with the notations of Lemma 3, Lemma 5 entails that there exists  $N \in \mathbb{N} \setminus \{0\}$  with

$$\left| \frac{1}{N} \mathbb{E}_{\Phi}(q_{0N}(\Phi')) - H(\Phi, \Phi') \right| < \varepsilon$$

so that

$$\frac{1}{N} \mathbb{E}_{\Phi}(q_{0N}(\Phi')) < H(\Phi, \Phi) - \varepsilon.$$

We then pick a neighborhood  $G$  of  $\Phi'$  in  $F_{\Phi_0}$  given by Lemma 3; in particular, for every subset  $G_{\Phi'}$  of  $G$  containing  $\Phi'$ ,

$$\mathbb{E}_{\Phi} \left| \ln \sup_{\varphi \in G_{\Phi'}} q_{0N}(\varphi) \right| < \infty.$$

Letting  $B_{1/t}$  be the open ball centered at  $\Phi'$  with radius  $1/t$ , the continuity of  $q_{0N}$  gives:

$$\ln \sup_{\varphi \in G \cap B_{1/t}} q_{0N}(\varphi) \xrightarrow[t \rightarrow \infty]{} \ln q_{0N}(\Phi').$$

Set now  $A_t = \left\{ \sup_{\varphi \in G \cap B_{1/t}} q_{0N}(\varphi) \leq 1 \right\}$ , and let  $A_t^c$  denote the complement of  $A_t$ . Notice that

$$\left| \ln \sup_{\varphi \in G \cap B_{1/t}} q_{0N}(\varphi) \right| = -\ln \left[ \sup_{\varphi \in G \cap B_{1/t}} q_{0N}(\varphi) \right] \mathbb{1}_{A_t} + \ln \left[ \sup_{\varphi \in G \cap B_{1/t}} q_{0N}(\varphi) \right] \mathbb{1}_{A_t^c}$$

which entails

$$\left| \ln \sup_{\varphi \in G \cap B_{1/t}} q_{0N}(\varphi) \right| \leq |\ln q_{0N}(\Phi')| + \left| \ln \sup_{\varphi \in G} q_{0N}(\varphi) \right|.$$

We can then use the dominated convergence theorem to get a neighborhood  $G_{\Phi'} \subset G$  of  $\Phi'$  in  $F_{\Phi_0}$  such that

$$\frac{1}{N} \mathbb{E}_{\Phi} \left| \ln \sup_{\varphi \in G_{\Phi'}} q_{0N}(\varphi) \right| \leq \frac{1}{N} \mathbb{E}_{\Phi} (\ln q_{0N}(\Phi')) + \frac{\varepsilon}{2} < H(\Phi, \Phi) - \frac{\varepsilon}{2}.$$

Now, because  $(Z_{st} = \ln \sup_{\varphi \in G_{\Phi'}} q_{st}(\varphi))$  is  $\mathbb{P}_{\Phi}$ -subadditive and ergodic, Kingman's theorem (see [22]) implies that there exists a finite constant  $H(\Phi, \Phi', G_{\Phi'})$  such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{\Phi} \left[ \ln \sup_{\varphi \in G_{\Phi'}} q_{0n}(\varphi) \right] = H(\Phi, \Phi', G_{\Phi'})$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \sup_{\varphi \in G_{\Phi'}} q_{0n}(\varphi) = H(\Phi, \Phi', G_{\Phi'}) \quad \mathbb{P}_{\Phi} - \text{a.s.}$$

Theorem 1.1 in [22] entails

$$H(\Phi, \Phi', G_{\Phi'}) \leq \frac{1}{N} \mathbb{E}_{\Phi} \left[ \ln \sup_{\varphi \in G_{\Phi'}} q_{0N}(\varphi) \right] < H(\Phi, \Phi) - \frac{\varepsilon}{2};$$

putting

$$p_{st}(\varphi | J(0) = j) = \mathcal{L}((Y_i, E_i, X_i)_{s+1 \leq i \leq t}, \varphi | J(0) = j)$$

and remarking that for all  $\varphi \in G_{\Phi'}$

$$\begin{aligned} q_{0n}(\varphi) &= \left( \sum_{i \in \mathcal{C}(\varphi)} \pi_i(\varphi) \right) \max_{i \in \mathcal{C}(\varphi)} p_{0n}(\varphi | J(0) = i) \\ &\geq \sum_{i \in \mathcal{C}(\varphi)} \pi_i(\varphi) p_{0n}(\varphi | J(0) = i) \\ &= p_{0n}(\varphi), \end{aligned}$$

one gets  $\ln \sup_{\varphi \in G_{\Phi'}} p_{0n}(\varphi) - \ln \sup_{\varphi \in G_{\Phi'}} q_{0n}(\varphi) \leq 0$  and thus

$$\limsup_{n \rightarrow \infty} \left\{ \frac{1}{n} \ln \sup_{\varphi \in G_{\Phi'}} p_{0n}(\varphi) \right\} \leq H(\Phi, \Phi', G_{\Phi'}) < H(\Phi, \Phi) - \frac{\varepsilon}{2}.$$

Cover now the compact set  $O^c \cap C$  by the  $G_{\Phi'_i}$ ,  $1 \leq i \leq d$ . We have

$$\sup_{\varphi \in O^c} \{ \ln p_{0n}(\varphi) - \ln p_{0n}(\Phi_0) \} \leq \max_{1 \leq i \leq d} \left\{ \ln \sup_{\varphi \in G_{\Phi'_i}} p_{0n}(\varphi) - \ln p_{0n}(\Phi_0) \right\} \xrightarrow{n \rightarrow \infty} -\infty$$

with  $\mathbb{P}_{\Phi_0}$ -probability 1. This shows that necessarily  $\widehat{\Phi}_p \in C$  a.s. for  $p$  large enough, and completes the proof. ■

Notice that since our model is not identifiable, any convergence result has to be stated modulo  $\sim$ . In that sense, this result is the best possible one.

Under some additional assumptions, one can apply the asymptotic normality theorem in [5] in order to obtain the one of our estimator. This result is rather technical: we refer the reader to [16] for details.

## 4 An EM algorithm to compute the MLE

We now give an EM algorithm, adapted from [35], allowing us to compute the MLE in our context. Recall the available data:

1. The values  $0 < t_1 < \dots < t_k = T$  of the  $\tau_i$ , *i.e.* the times when  $N$  jumps (equivalently, the inter-event times  $y_1, \dots, y_k$ , where  $y_j = t_j - t_{j-1}$ ,  $t_0 = 0$ );
2.  $e_1, \dots, e_k$  the successive values of the  $E_k$ ;
3.  $x_1, \dots, x_k$  the successive values of the jumps of  $S$ .

We want to estimate

1. The elements  $\ell_{ij}$  of the transition intensity matrix  $L$  of  $J$ ;
2. The jump intensities  $\lambda_i$  of  $N$ ;
3. The probabilities  $p_i(e)$ , where  $e$  is a subset of  $\{1, \dots, n\}$ ;
4. The parameters  $\theta(i, e)$ .

We let  $0 < u_1 < \dots < u_m < T$  be the jump times of  $J$  in the time interval  $[0, T]$ ,  $u_0 = 0$  and  $u_{m+1} = T$ ; let further  $s_i$  be the state of  $J$  on the interval  $[u_{i-1}, u_i[$ ,  $\Delta u_i = u_i - u_{i-1}$  and  $z_i$  be the number of jumps of  $N$  in the interval  $[u_{i-1}, u_i[$ .

Recall that, if  $N'$  is an homogeneous Poisson process, then given  $\{N'(t) = n\}$ , the event times of  $N'$  in the interval  $[0, t]$  are uniformly distributed. Consequently, Bayes' formula implies that the complete likelihood of the data is

$$\begin{aligned} \mathcal{L}^c &= \pi_{s_1} \left[ \prod_{i=1}^m \frac{\ell_{s_i, s_{i+1}}}{-\ell_{s_i, s_i}} \cdot (-\ell_{s_i, s_i} \exp(\ell_{s_i, s_i} \Delta u_i)) \right] \exp(\ell_{s_{m+1}, s_{m+1}} \Delta u_{m+1}) \\ &\times \left[ \prod_{i=1}^{m+1} \frac{(\lambda_{s_i} \Delta u_i)^{z_i}}{z_i!} \exp(-\lambda_{s_i} \Delta u_i) \cdot \frac{z_i!}{(\Delta u_i)^{z_i}} \right] \\ &\times \prod_{i=1}^r \left[ \prod_{\substack{e \subset \{1, \dots, n\} \\ e \neq \emptyset}} p_i(e)^{\text{card}(A_i(e))} \prod_{j \in A_i(e)} \mathbb{P}_{\theta(i, e)}(\forall m \in e, X_m = x_{m, j}) \right] \cdot p_i(\emptyset)^{\text{card}(A_i(\emptyset))} \end{aligned}$$



where  $A_i(e) = \{j \in \{1, \dots, k\} \mid J(t_j) = i, e_j = e\}$  stands for the set of the jump times of  $N$  when the  $S_k$  with  $k \in e$  (and only these) jump and  $J$  is in state  $i$ ;  $A_i(\emptyset)$  stands for the set of the jump times of  $N$  when none of the  $S_k$  jumps and  $J$  is in state  $i$ .

From that identity, we deduce that the complete log-likelihood is

$$\begin{aligned} \ln \mathcal{L}^c &= \sum_{i=1}^r \mathbb{1}_{\{X(0)=i\}} \ln(\pi_i) + \sum_{i=1}^r T_i \ell_{ii} + \sum_{i=1}^r \sum_{\substack{j=1 \\ j \neq i}}^r m_{ij}(T) \ln(\ell_{ij}) + \sum_{i=1}^r (n_i \ln(\lambda_i) - \lambda_i T_i) \\ &\quad + \sum_{i=1}^r \sum_{e \subset \{1, \dots, n\}} \text{card}(A_i(e)) \ln(p_i(e)) \\ &\quad + \sum_{i=1}^r \sum_{\substack{e \subset \{1, \dots, n\} \\ e \neq \emptyset}} \sum_{j=1}^k \ln \mathbb{P}_{\theta(i, e)}(\forall m \in e, X_m = x_{m, j}) \mathbb{1}_{\{j \in A_i(e)\}} \end{aligned}$$

where

1.  $T_i = \int_0^T \mathbb{1}_{\{J(u)=i\}} du$  is the time spent by the process  $J$  in state  $i$  until time  $T$ ;
2.  $m_{ij}(T) = \text{card}(\{s : 0 < s \leq T \mid J(s_-) = i, J(s) = j\})$  is the number of jumps from state  $i$  to state  $j$  of the process  $J$ ;
3.  $n_i = \sum_{j=1}^k \mathbb{1}_{\{J(t_j)=i\}}$  is the number of events that occurred when  $J$  is in state  $i$ .

**The M step.** We now compute the conditional expectation of  $\ln \mathcal{L}^c(\Phi)$  under a parameter  $\varphi$ , given the event  $\{N(u), S(u), 0 \leq u \leq T\}$ : one has

$$\begin{aligned} &\mathbb{E}_\varphi(\ln \mathcal{L}^c(\Phi) \mid N(u), S(u), 0 \leq u \leq T) \\ &= \sum_{i=1}^r \widehat{\mathbb{1}_{\{X(0)=i\}}} \ln(\pi_i) + \sum_{i=1}^r \widehat{T_i} \ell_{ii} + \sum_{i=1}^r \sum_{\substack{j=1 \\ j \neq i}}^r \widehat{m_{ij}(T)} \ln(\ell_{ij}) + \sum_{i=1}^r (\widehat{n_i} \ln(\lambda_i) - \lambda_i \widehat{T_i}) \\ &\quad + \sum_{i=1}^r \sum_{e \subset \{1, \dots, n\}} \widehat{\text{card}(A_i(e))} \ln(p_i(e)) + \sum_{i=1}^r \sum_{\substack{e \subset \{1, \dots, n\} \\ e \neq \emptyset}} \sum_{j=1}^k \ln \mathbb{P}_{\theta(i, e)}(\forall m \in e, X_m = x_{m, j}) \widehat{\mathbb{1}_{\{j \in A_i(e)\}}} \end{aligned}$$

where  $\widehat{A} = \mathbb{E}_\varphi(A \mid N(u), S(u), 0 \leq u \leq T)$ .

For  $T$  large enough, the first term may be neglected; recalling that

$$\ell_{ii} = - \sum_{\substack{j=1 \\ j \neq i}}^r \ell_{ij}, \quad p_i(\emptyset) = 1 - \sum_{\substack{e \subset \{1, \dots, n\} \\ e \neq \emptyset}} p_i(e), \quad \widehat{\text{card}(A_i(\emptyset))} = \widehat{n_i} - \sum_{\substack{e \subset \{1, \dots, n\} \\ e \neq \emptyset}} \widehat{\text{card}(A_i(e))},$$

one gets, for all  $i, j \in \{1, \dots, r\}$  and  $i \neq j$ , the identities

$$\begin{aligned} \widehat{p_i}(e) &= \frac{\widehat{\text{card}(A_i(e))}}{\widehat{n_i}}, \quad \widehat{\ell_{ij}} = \frac{\widehat{m_{ij}(T)}}{\widehat{T_i}}, \quad \widehat{\lambda_i} = \frac{\widehat{n_i}}{\widehat{T_i}}, \\ \sum_{j=1}^k \frac{\partial}{\partial \theta(i, e)} \ln \mathbb{P}_{\theta(i, e)}(\forall m \in e, X_m = x_{m, j}) \Big|_{\theta(i, e) = \widehat{\theta}(i, e)} \widehat{\mathbb{1}_{\{j \in A_i(e)\}}} &= 0, \end{aligned}$$

where  $\widehat{p}_i(e)$ ,  $\widehat{\ell}_{ij}$  and  $\widehat{\lambda}_i$  are the desired estimators, and the last set of equations is to be solved taking the properties of the statistical model  $(\mathbb{P}_\theta)$  into account.

**The E step.** According to Lemma 9, if  $A(e) = \bigcup_{i=1}^r A_i(e) = \{j \in \{1, \dots, k\} \mid e_j = e\}$ , then

$$\begin{aligned}\widehat{T}_i &= \int_0^T \frac{\mathbb{P}_\varphi(J(v) = i, N(u), S(u), 0 \leq u < v)}{\mathbb{P}_\varphi(N(u), S(u), 0 \leq u \leq T)} \\ &\quad \times \mathbb{P}_\varphi(N(u), S(u), v \leq u \leq T \mid J(v) = i) dv, \\ \widehat{n}_i &= \sum_{q=1}^k \frac{\mathbb{P}_\varphi(J(t_q) = i, N(u), S(u), 0 \leq u \leq T)}{\mathbb{P}_\varphi(N(u), S(u), 0 \leq u \leq T)}, \\ \mathbb{1}_{\widehat{\{j \in A_i(e)\}}} &= \mathbb{1}_{\{j \in A(e)\}} \mathbb{P}_\varphi(J(t_j) = i \mid N(u), S(u), 0 \leq u \leq T), \\ \text{card}(\widehat{A_i(e)}) &= \sum_{j=1}^k \mathbb{1}_{\widehat{\{j \in A_i(e)\}}} = \sum_{j=1}^k \mathbb{1}_{\{j \in A(e)\}} \mathbb{P}_\varphi(J(t_j) = i \mid N(u), S(u), 0 \leq u \leq T), \\ \widehat{m_{ij}(T)} &= \ell_{ij}(\varphi) \int_0^T \frac{\mathbb{P}_\varphi(J(v) = i, N(u), S(u), 0 \leq u < v)}{\mathbb{P}_\varphi(N(u), S(u), 0 \leq u \leq T)} \\ &\quad \times \mathbb{P}_\varphi(N(u), S(u), v \leq u \leq T \mid J(v) = j) dv.\end{aligned}$$

Let  $w_i$  be the column vector of size  $r$  with all entries except the  $i$ -th equal to 0, and its  $i$ -th entry equal to 1. Firstly,

$$\mathbb{P}_\varphi(N(u), S(u), 0 \leq u < v, J(v) = i) = \pi(\varphi) \left( \prod_{q=1}^{N(v)} g(y_q, e_q, x_q, \varphi) \right) \overline{F}(v - t_{N(v)}, \varphi) w_i.$$

Secondly, if  $w_i^t$  is the transpose of  $w_i$ ,

$$\begin{aligned}\mathbb{P}(N(u), S(u), v \leq u \leq T, \varphi \mid J(v) = i) \\ = w_i^t g(t_{N(v)+1} - v, e_{N(v)+1}, x_{N(v)+1}, \varphi) \left( \prod_{q=N(v)+2}^k g(y_q, e_q, x_q, \varphi) \right) \eta,\end{aligned}$$

and finally

$$\begin{aligned}\mathbb{P}_\varphi(J(t_q) = i, N(u), S(u), 0 \leq u \leq T) \\ = \pi(\varphi) \left( \prod_{p=1}^q g(y_p, e_p, x_p, \varphi) \right) w_i w_i^t \left( \prod_{p=q+1}^k g(y_p, e_p, x_p, \varphi) \right) \eta.\end{aligned}$$

$\theta$  is generally estimated with a numerical (e.g. quasi-Newton) method.

**Procedure.** Here, we describe a way to implement our algorithm, by induction on  $\ell \in \mathbb{N}$ .

Define, if  $\Phi_\ell$  is the parameter estimate at step  $\ell$ ,

1.  $G_\ell(0) = \pi(\Phi_\ell)$  and  $\forall 0 \leq q \leq k-1$ ,  $G_\ell(q+1) = G_\ell(q) \cdot g(y_{q+1}, e_{q+1}, x_{q+1}, \Phi_\ell)$ ;
2.  $D_\ell(k) = \eta$  and  $\forall 0 \leq q \leq k-1$ ,  $D_\ell(k-q-1) = g(y_{k-q}, e_{k-q}, x_{k-q}, \Phi_\ell) \cdot D_\ell(k-q)$ .

Set then  $A_{ij}(\Phi_\ell) = B_i(\cdot, \Phi_\ell) = C_i(\Phi_\ell) = 0$  and do, for all  $q \in \mathbb{N}$  such that  $1 \leq q \leq k$ ,

$$\begin{aligned} A_{ij}(\Phi_\ell) &\leftarrow A_{ij}(\Phi_\ell) + \int_{t_{q-1}}^{t_q} G_\ell(q-1) \overline{F}(t - t_{q-1}, \Phi_\ell) w_i w_j^t g(t_q - t, e_q, x_q, \Phi_\ell) D_\ell(q) dt, \\ B_i(q, \Phi_\ell) &\leftarrow G_\ell(q) w_i w_i^t D_\ell(q), \\ C_i(\Phi_\ell) &\leftarrow C_i(\Phi_\ell) + B_i(q, \Phi_\ell). \end{aligned}$$

The estimates at step  $\ell + 1$  are then

$$\hat{p}_i(e) = \frac{\sum_{j=1}^k \mathbb{1}_{\{j \in A(e)\}} B_i(j, \Phi_\ell)}{C_i(\Phi_\ell)}, \quad \hat{\ell}_{ij} = \ell_{ij}(\Phi_\ell) \cdot \frac{A_{ij}(\Phi_\ell)}{A_{ii}(\Phi_\ell)}, \quad \hat{\lambda}_i = \frac{C_i(\Phi_\ell)}{A_{ii}(\Phi_\ell)},$$

and the  $\hat{\theta}(i, e)$  that maximize the functionals

$$\theta \mapsto \sum_{j=1}^k \ln \mathbb{P}_\theta(\forall m \in e, X_m = x_{m,j}) B_i(j, \Phi_\ell) \mathbb{1}_{\{j \in A(e)\}}.$$

## 5 A posteriori reconstruction of the states, with a maximum likelihood method

Once the parameters of the model are estimated, it can be interesting to estimate the successive states of the Markov chain  $(J_i)$ . To this end, we can adapt the procedure described in [28]: consider the log-likelihood of both the observed and missing data

$$(j_0, \dots, j_k) \mapsto \ln(\pi_{j_0}(\hat{\Phi})) + \sum_{i=1}^k \ln g_{j_{i-1}, j_i}(y_i, e_i, x_i, \hat{\Phi}).$$

An estimator of  $(j_0, \dots, j_k)$  is then a  $(k+1)$ -tuple  $(\hat{j}_0, \dots, \hat{j}_k)$  which maximizes this functional. Such an estimator has excellent properties, see [8]. From a practical point of view, one may reconstruct the states using the Viterbi algorithm (see [37]), namely:

1. Set  $V_j = 0$  and  $C_j = [j]$  for all  $j \in \{1, \dots, r\}$ , and  $q = 1$ .
2. If  $q \geq k+1$ , go to step 6. Otherwise, set

$$\alpha_{i,j}^{(q)} = \ln g_{ij}(y_{k-q+1}, e_{k-q+1}, x_{k-q+1}, \hat{\Phi}).$$

3. For all  $i, j \in \{1, \dots, r\}$ , compute  $\beta_{i,j}^{(q)} = \alpha_{i,j}^{(q)} + V_j$  and an index  $j_i^{(q)}$  such that  $\beta_{i,j_i^{(q)}}^{(q)} = \max_{j \in \{1, \dots, r\}} \beta_{i,j}^{(q)}$ .
4. For all  $i \in \{1, \dots, r\}$ , replace  $V_i$  by  $\beta_{i,j_i^{(q)}}^{(q)}$  and  $C_i$  by  $[j_i^{(q)}, C_i]$ .
5. Replace  $q$  by  $q+1$  and go back to step 2.
6. Find an index  $i$  such that  $V_i = \max_{j \in \{1, \dots, r\}} V_j$ .

An estimate of the states is then the sequence  $(\hat{j}_0, \dots, \hat{j}_k) = C_i$ .

## 6 Numerical illustrations

### 6.1 Computing a first estimate

Providing a first estimate for an iterative algorithm is usually a daunting task. Here, we describe a procedure, adapted from the one described in [28], that worked quite well in our examples:

1. Compute the average of the inter-event times  $\widehat{\lambda}^* = k/T$ , and mobile averages of the inter-event times  $y_i$ , denoted by  $z_i$  (for the first and last times of the observed sample, put  $z_i = y_i$ ).
2. Set  $\widehat{J}(\cdot) = 0$ ; pick  $q_1 \leq 1 < q_2 < \dots < q_{r-1}$ . For all  $i \in \{1, \dots, k\}$ :
  - (a) if  $z_i > 1/(q_1 \widehat{\lambda}^*)$ , set  $\widehat{J}(t_i) = 1$ ;
  - (b) for all  $j \in \{1, \dots, r-2\}$ , if  $1/(q_{j+1} \widehat{\lambda}^*) < z_i \leq 1/(q_j \widehat{\lambda}^*)$ , set  $\widehat{J}(t_i) = j+1$ ;
  - (c) if  $z_i \leq 1/(q_{r-1} \widehat{\lambda}^*)$ , set  $\widehat{J}(t_i) = r$ .
3. Compute  $\widehat{n}_j = \sum_{i=1}^{k-1} \mathbb{1}_{\{\widehat{J}(t_i)=j\}}$  for  $j \in \{1, \dots, r\}$ .
4. Compute, for all  $i, j \in \{1, \dots, r\}$

$$\widehat{P}_{ij} = \frac{\sum_{\ell=2}^k \mathbb{1}_{\{\widehat{J}(t_{\ell-1})=i, \widehat{J}(t_\ell)=j\}}}{\widehat{n}_i},$$

which is the first estimate of  $P_{ij}$ , the probability that the Markov chain  $(J(t_k))_{k \geq 0}$  jumps from state  $i$  to state  $j$ .

5. Calculate, for all  $j \in \{1, \dots, r\}$ ,  $\widehat{\pi}_j = \frac{\widehat{n}_j + \mathbb{1}_{\{\widehat{J}(t_k)=j\}}}{k}$ , the first estimate of  $\pi_j$ .
6. Thanks to the identities

$$\forall j \in \{1, \dots, r\} \quad \lambda_j = \lambda^* \pi_j a_j^{-1} \quad \text{and} \quad L = \Lambda(\text{Id} - P^{-1}),$$

(where  $\lambda^* = \sum_{j=1}^r \lambda_j a_j$  is the average jump rate of  $N$ ), consider  $L$  and  $\Lambda$  as functions of  $a_1, \dots, a_{r-1}$ , and maximize the complete likelihood with respect to the parameters  $a_1, \dots, a_{r-1}$  given  $\widehat{\lambda}^*, \widehat{\pi}_1, \dots, \widehat{\pi}_r, \widehat{P}, y_1, \dots, y_k$  and  $\widehat{J}$ : let  $\widehat{a}_1, \dots, \widehat{a}_{r-1}$  be the estimate obtained this way.

7. For all  $j \in \{1, \dots, r\}$ , compute  $\widehat{\lambda}_j = \widehat{\lambda}^* \widehat{\pi}_j \widehat{a}_j^{-1}$ , let  $\widehat{\Lambda}$  be the diagonal matrix with coefficients  $\widehat{\lambda}_1, \dots, \widehat{\lambda}_r$  in that order and compute  $\widehat{L} = \widehat{\Lambda}(\text{Id} - \widehat{P}^{-1})$ . These are rough estimates for  $\Lambda$  and  $L$ .

8. Use  $\widehat{L}$  and  $\widehat{\Lambda}$  as initial values for an EM algorithm to provide estimates for  $L$  and  $\Lambda$  (see [35]), which we denote by  $\overline{L}$  and  $\overline{\Lambda}$ . Compute the corresponding stationary distributions  $\overline{a}$  and  $\overline{\pi}$ .
9. Perform a state reconstruction of  $J$  with the Viterbi algorithm using  $\overline{L}$  and  $\overline{\Lambda}$ , and let  $\overline{J}$  be the process obtained this way.
10. For all  $j \in \{1, \dots, r\}$ , calculate  $\overline{n}_j = \sum_{i=1}^{k-1} \mathbb{1}_{\{\overline{J}(t_i)=j\}}$ .
11. For all  $i_1, \dots, i_n \in \{0, 1\}$  and  $j \in \{1, \dots, r\}$ , if  $e$  is the subset of  $\{1, \dots, n\}$  such that  $k \in e \Leftrightarrow i_k = 1$ , compute

$$\overline{p}_j(e) = \frac{1}{\overline{n}_j} \sum_{\ell=1}^{k-1} \mathbb{1}_{\{\overline{J}(t_\ell)=j\}} \mathbb{1}_{\{\forall p \in \{1, \dots, n\}, S_p(t_\ell) - S_p(t_{\ell-1}) > 0 \Leftrightarrow i_s = 1\}}$$

which is the initial estimate of  $p_j(e)$ .

12. For all  $j = 1, \dots, r$  and  $e \neq \emptyset$ , consider the  $X_i$  such that  $\overline{J}(t_i) = j$  and  $E_i = e$  as independent and identically distributed random variables with parameter  $\theta(j, e)$ , and estimate  $\theta(j, e)$  with a standard method (maximum likelihood method for instance).

This procedure is adapted in the particular case when  $\lambda_1 < \dots < \lambda_r$  strongly differ, which shall be the case in our numerical study below.

## 6.2 A non-life insurance example

We now use our algorithm on a real set of non-life insurance data. From January 2004 to November 2009, 594 accidents corresponding to blazes causing industrial damages or losses were observed. The days of these events were recorded, and so were, if necessary, the compensations for the victims; the processes  $N$  and  $S$  obtained this way are shown on Figure 1–2. This situation corresponds to the case  $\alpha = n = 1$  of our model. We finally choose  $r = 2$ , which is justified by the fact that the MLE, computed only for  $L$  and  $\Lambda$  with  $r = 3$  sets all parameters corresponding to the third state to 0. Before modeling the claims themselves, the parameters of this model are

1.  $\ell_{12}$  and  $\ell_{21}$ , the jump rates of the hidden Markov process  $J$ ;
2.  $\lambda_1$  and  $\lambda_2$ , the jump intensities of the shock counting process  $N$ ;
3.  $p_1(1)$  and  $p_2(1)$ , the probabilities that, when an accident happens, the insurance firm has to compensate.

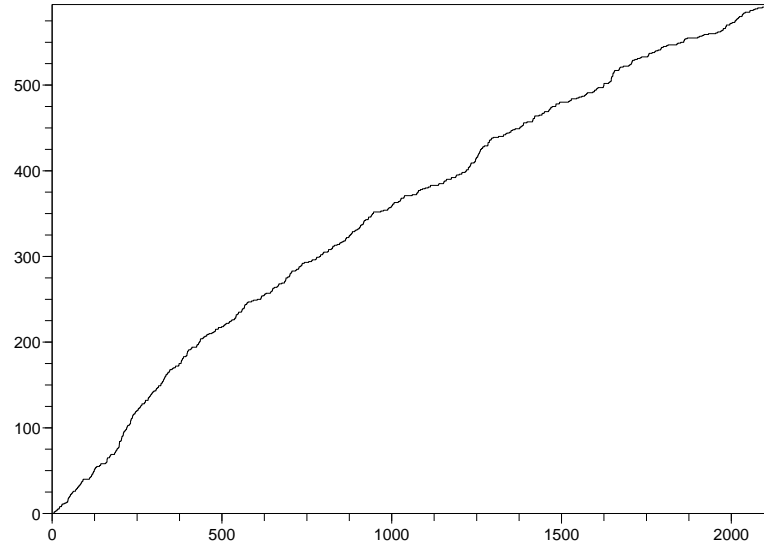


Figure 1: The counting process  $N$

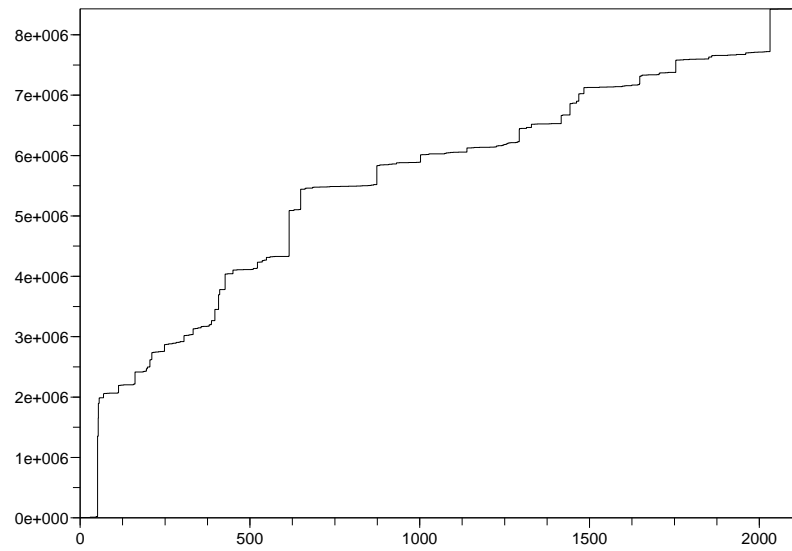


Figure 2: The loss process  $S$

As for the claim sizes, a quick analysis of the data shows that some claims have a small size and a few others are very large, which prevents us from modeling the situation by a log-Normal, Gamma or Generalized Pareto distribution (GPD). In actuarial statistics, one may

either try to separate so-called attritional claims and large claims thanks to some threshold as in many Solvency II partial internal models, or deal directly with a mixture of distributions, or with a distribution that looks like Lognormal or Gamma distributions for small values and gets more and more Pareto-type for large values, like the Champervowne distribution (see [9, 10] and [20]). Another possibility is to use a classical kernel density estimator after transforming the data (see [6]). Here, we use a mixture of a light-tailed and a heavy-tailed distribution, namely a Gamma distribution and a GPD.  $\mathbb{P}_\theta$  then has density

$$x \mapsto q \frac{(bx)^{a-1}}{\Gamma(a)} b e^{-bx} \mathbb{1}_{\{x>0\}} + (1-q) \frac{1}{\sigma} \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1-1/\xi} \mathbb{1}_{\{x>\mu\}}$$

where  $a, b, \sigma, \xi > 0$ ,  $0 < q < 1$  and  $\mu = 49.33$  is the minimal (observed) claim size (the unit is the euro).

Consequently, the parameters to be estimated are  $\ell_{12}, \ell_{21}, \lambda_1, \lambda_2, p_1(1), p_2(1), a_1, a_2, b_1, b_2, \sigma_1, \sigma_2, \xi_1, \xi_2, q_1$  and  $q_2$ .

Estimating the parameters via the EM algorithm, with a quasi-Newton algorithm to estimate the parameters  $a_i, b_i, \sigma_i, \xi_i$  and  $q_i$  during the M step gives the following results:

$$\begin{aligned} \hat{L} &= \begin{pmatrix} -0.0065 & 0.0065 \\ 0.0018 & -0.0018 \end{pmatrix}, \quad \hat{\Lambda} = \begin{pmatrix} 0.462 & 0 \\ 0 & 0.214 \end{pmatrix}, \\ \hat{p}(1) &= \begin{pmatrix} 0.963 & 0 \\ 0 & 0.947 \end{pmatrix}, \quad \hat{p}(0) = \begin{pmatrix} 0.037 & 0 \\ 0 & 0.053 \end{pmatrix}, \\ \hat{a} &= \begin{pmatrix} 4.52 \\ 4.14 \end{pmatrix}, \quad \hat{b} = \begin{pmatrix} 0.011 \\ 0.0073 \end{pmatrix}, \quad \hat{\sigma} = \begin{pmatrix} 1145 \\ 1216 \end{pmatrix}, \quad \hat{\xi} = \begin{pmatrix} 1.45 \\ 1.31 \end{pmatrix}, \quad \hat{q} = \begin{pmatrix} 0.230 \\ 0.335 \end{pmatrix}. \end{aligned}$$

The claim sizes thus have infinite means in both states in theory. This means that the tail of the claim size distribution is very heavy. However, reinsurance mechanisms and other guarantees may enable the insurer to provide insurance coverage of those risks up to some high threshold level. A further analysis then shows that

1. Sojourn times in state 1 are on average 3.5 times shorter than in state 2;
2. There are more accidents when  $J$  is in state 1 than in state 2;
3. Because  $\hat{p}_1(1)$  is slightly greater than  $\hat{p}_2(1)$ , these accidents cause more losses to the insurance firm;
4. Losses in state 1 are more likely to be heavy-tailed than in state 2.

An *a posteriori* reconstruction of the states of  $J$  is given in Figure 3.

### 6.3 A life insurance data set

Let us now present an application in the life insurance field. From January 2006 to July 2010, 1507 closures of savings accounts (also called surrenders) were observed. The months

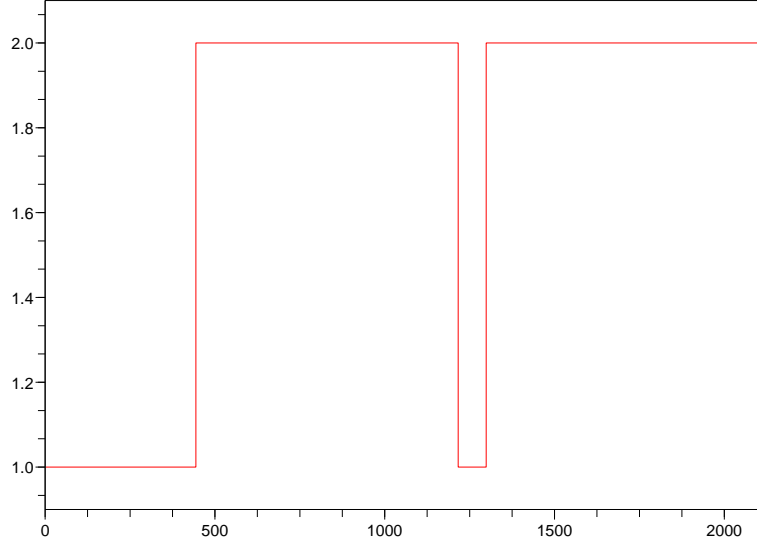


Figure 3: *A posteriori* reconstruction of the states of  $J$

of these events were recorded, along with the amount of money withdrawn. Early surrenders can be regarded as claims for the insurance company in some cases, because it corresponds to a drop in future business, and because sometimes the insurer has been unable to charge all the fees (that are often partly paid by the policyholder at each time period and not upfront) before the surrender. Surrender risk is complex: tax and penalty relief, interest rate levels, competition between insurance companies, as well as other factors are at stake. For a review on surrender triggers, the interested reader might consult [29] or [26]. In the present data study, we are interested in the big picture in a quite stable regime (and not in prediction of future surrender rates): in the considered period, the portfolio seems to have been pretty stable, mainly sensitive to external competition (which is difficult to observe in practice). We assume that conditionally with respect to the state of the environment, the probability for one policyholder to surrender her contract does not depend on the amount of savings. To set a precise date for the  $k$ -th surrender, we draw a uniform random variable and add it to the month of this event to obtain an exact date. Here, the “claims” are the amounts of money withdrawn; the processes  $N$  and  $S$  are represented on Figure 4–5. Again, this situation fits the case  $\alpha = n = 1$  of our model; we use a two-state model for this situation, so that the parameters are

1.  $\ell_{12}$  and  $\ell_{21}$ , the jump rates of the hidden Markov process  $J$ ;
2.  $\lambda_1$  and  $\lambda_2$ , the jump intensities of the shock counting process  $N$ .



Note that in this example, there is no need to estimate  $p_1(1)$  and  $p_2(1)$ . On the graphs below, the unit of time is the month:

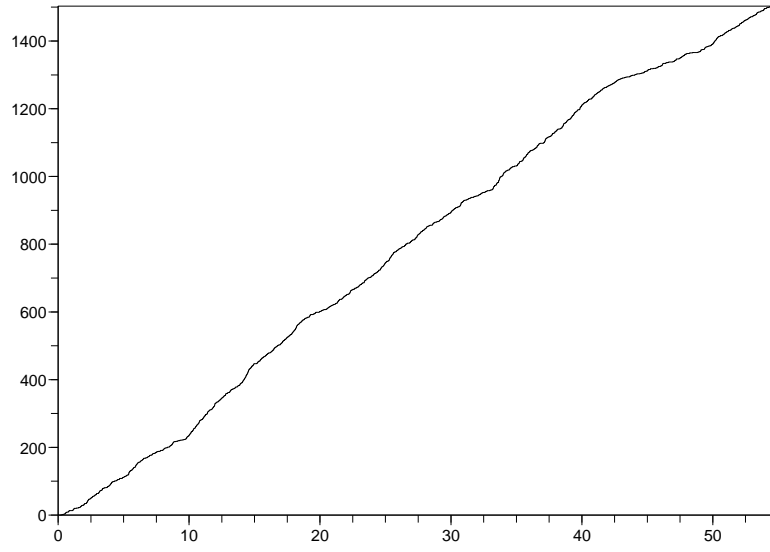


Figure 4: The counting process  $N$

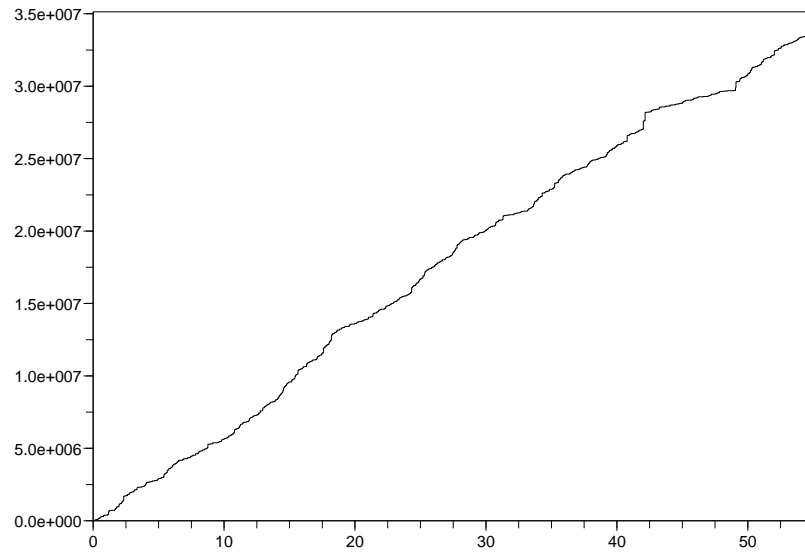


Figure 5: The process  $S$  representing the cumulative amount of money withdrawn

In state 1, we use a mixture of a light-tailed and a heavy-tailed distribution, namely a Weibull distribution and a GPD, the density of  $\mathbb{P}_\theta$  then being

$$x \mapsto q \frac{a}{b} \left[ \frac{x - \mu}{b} \right]^{a-1} e^{-((x-\mu)/b)^a} \mathbb{1}_{\{x > \mu\}} + (1 - q) \frac{1}{\sigma} \left( 1 + \frac{\xi(x - \mu)}{\sigma} \right)^{-1-1/\xi} \mathbb{1}_{\{x > \mu\}}$$

where  $a, b, \sigma, \xi > 0$ ,  $0 < q < 1$  and  $\mu = 1.1$  is the minimal (observed) amount (the unit is the euro). In state 2, we fit a GPD, whose density is

$$x \mapsto \frac{1}{\sigma} \left( 1 + \frac{\xi(x - \mu)}{\sigma} \right)^{-1-1/\xi} \mathbb{1}_{\{x > \mu\}} \quad (1)$$

where  $\mu, \sigma, \xi > 0$ . Of course, surrender amounts are not completely independent at the microscopic level as each policyholder has a certain balance on his savings account that is known at a precise date. We are aware that in theory, the  $X_i$  are not independent and identically distributed in each state, but in practice there are enough policyholders and enough randomness in the surrendered amounts for this assumption to be acceptable in practice at the macroscopic level in each state of the environment (this is supported by statistical tests).

Consequently, the parameters to be estimated are  $\ell_{12}, \ell_{21}, \lambda_1, \lambda_2, a, b, \sigma_1, \sigma_2, \xi_1, \xi_2$  and  $q$ .

Estimating the parameters via the EM algorithm, with a quasi-Newton algorithm to estimate the parameters  $a, b, \sigma_i, \xi_i$  and  $q$  during the M step gives the following results:

$$\begin{aligned} \hat{L} &= \begin{pmatrix} -0.254 & 0.254 \\ 0.373 & -0.373 \end{pmatrix}, \quad \hat{\Lambda} = \begin{pmatrix} 34.2 & 0 \\ 0 & 17.4 \end{pmatrix}, \\ \hat{a} &= 1.65, \quad \hat{b} = 9141, \quad \hat{\sigma} = \begin{pmatrix} 22350 \\ 14591 \end{pmatrix}, \quad \hat{\xi} = \begin{pmatrix} 0.17 \\ 0.40 \end{pmatrix}, \quad \hat{q} = 0.306. \end{aligned}$$

An *a posteriori* reconstruction of the states of  $J$  is shown in Figure 6. Note that results show that during some fierce competition periods, surrender rates become more important (they double from one state to the other). In the state where surrender rates are higher, the surrendered amount fitted distribution is composed of a light-tailed part and a heavy-tailed part, whereas for smaller surrender rates, this distribution does not incorporate any light-tailed part. This suggests that policies with smaller facial amounts are more sensitive to changes in the environment. Once again, here, the heavy-tailed part must be regarded as a statistical fit, and the tail would have to be cut at an appropriate level *a posteriori*.

## 6.4 Simulations in the multivariate setting

### 6.4.1 Motivation

One of the main purposes of insurance is risk diversification and mutualization: the law of large numbers and the central limit theorem often apply in practice when independence between individual risks is not too unrealistic. For example, this works quite well for motor

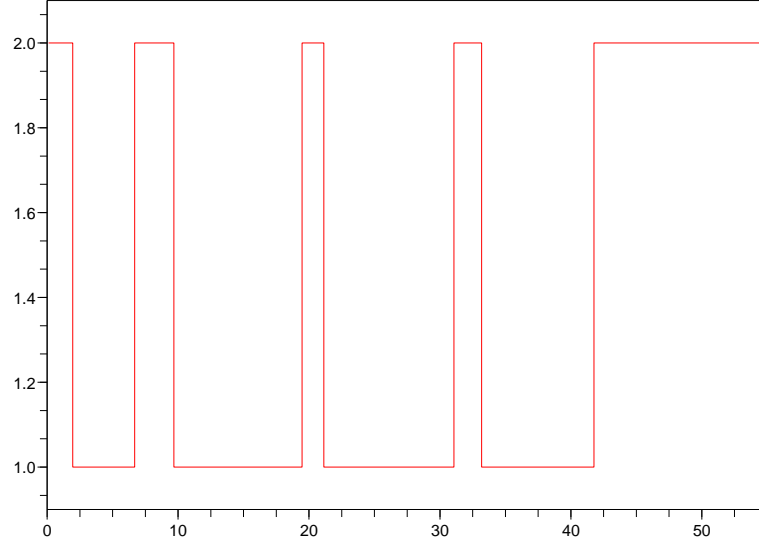


Figure 6: *A posteriori* reconstruction of the states of  $J$

insurance portfolios (without motor liability insurance) at the national level. However, when it comes to hurricane risks or earthquake risks, individual risks are only conditionally independent with respect to the occurrence or not of such events in the country. This correlation makes it difficult to diversify those risks at the national level, and one often uses reinsurance: risks are then diversified at the global level (floods in Australia, tsunamis in Asia, hurricanes in the East Coast of North America, earthquakes in Japan, Monte Carlo and San Francisco, storms in Europe for instance). Nevertheless, those risks are not really independent, as some (often ignored) correlation factors are present. Even if they are geographically scattered, meteorological phenomena like the El Nino-La Nina Southern Oscillation (ENSO) may simultaneously influence claim occurrence and severity in those different zones. For example, it is now accepted that the probabilities of severe floods in Australia, strong snowstorms in North America and hurricanes on the US East Coast increase during La Nina episodes, while other kinds of events are more likely during El Nino episodes. To build a model for ENSO and to understand all its impacts on different areas of the world is far beyond the scope of this paper. Of course, ENSO is observed and can be (partly) measured, its behavior is not really Markovian and claim arrival processes feature seasonality. There are certainly other kinds of unobserved environment processes that jointly modulate claim processes in different regions of the world. In our illustrative example, we just imagine that some unobserved Markov process influences claim frequencies in three regions A ( $k = 1$ ), B ( $k = 2$ ) and C ( $k = 3$ ). Regions A and B are assumed to be close to each other, so that common shocks

(events that simultaneously cause claims in both regions) are possible. In our example, phase changes are more frequent than for the ENSO cycle. We simulate the corresponding multivariate risk process, and we check whether it would be possible or not for us to estimate the parameters of the model and to re-build the states of the environment modulating process (without observing it of course).

#### 6.4.2 A model with 2 states of the environment

We first assume that  $r = 2$ : in state 1, claims are less frequent and less severe in the three zones, and common shocks are not present ( $p_1(e) = 0$  if  $\text{Card}(e) \geq 2$ ). In state 2, claims are more likely and more severe in average, and common shocks are possible for zones A and B ( $p_2(\{1, 2\}) > 0$ ). Take  $\lambda_1 = 20$ ,  $\lambda_2 = 200$ ,  $p_1(\{1\}) = p_1(\{2\}) = 0.3$ ,  $p_1(\{3\}) = 0.4$ ,  $p_2(\{1\}) = p_2(\{2\}) = 0.2$ ,  $p_2(\{3\}) = 0.4$  and  $p_2(\{1, 2\}) = 0.2$ . The univariate claim severity distributions are chosen to be GP distributed as in (1), with the parameters being

$$\begin{aligned}\mu(\{1\}) &= \mu(\{2\}) = \mu(\{3\}) = 1, \\ \sigma(1, \{1\}) &= \sigma(1, \{2\}) = \sigma(1, \{3\}) = 1, \\ \sigma(2, \{1\}) &= \sigma(2, \{2\}) = \sigma(2, \{3\}) = 20, \\ \xi(1, \{1\}) &= \xi(1, \{2\}) = \xi(1, \{3\}) = 1/2, \\ \xi(2, \{1\}) &= \xi(2, \{2\}) = \xi(2, \{3\}) = 2.\end{aligned}$$

Univariate claims are therefore more severe in average and in the tail for state 2 for all three lines. As far as the bivariate claims in state 2 are concerned, we model them by a bivariate GPD as in [7, 11]; namely, their density has the form

$$(x, y) \mapsto \frac{\alpha(\alpha + 1)}{\sigma_1 \sigma_2} \left( 1 + \frac{x - \mu_1}{\sigma_1} + \frac{x - \mu_2}{\sigma_2} \right)^{-\alpha-2} \mathbb{1}_{\{x > \mu_1\}} \mathbb{1}_{\{y > \mu_2\}}$$

where  $\alpha, \mu_1, \mu_2, \sigma_1, \sigma_2 > 0$ , and we choose

$$\mu(\{1, 2\}) = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \quad \sigma(2, \{1, 2\}) = \begin{pmatrix} 30 \\ 20 \end{pmatrix}, \quad \alpha(2, \{1, 2\}) = 2.$$

Assume that we observe the multivariate claim process during 30 years, and that the average time spent in state 1 (before switching to state 2) is 1 year, while the average time spent in state 2 (before switching to state 1) is 3 months. Namely,  $\ell_{12} = 1$  and  $\ell_{21} = 4$ .

The estimate of  $\mu(\{e\})$ ,  $e \neq \emptyset$  is chosen as the vector of the minima of the claims arising when a shock affects simultaneously the lines  $L_{k_1}, \dots, L_{k_p}$ , with  $e = \{k_1, \dots, k_p\}$ . Results are given below:

$$\begin{aligned}\widehat{\ell}_{12} &= 1.064, \quad \widehat{\ell}_{21} = 3.891, \\ \widehat{\lambda}_1 &= 21.21, \quad \widehat{\lambda}_2 = 195.7, \\ \widehat{p}_1(\{1\}) &= 0.340, \quad \widehat{p}_1(\{2\}) = 0.276, \quad \widehat{p}_1(\{3\}) = 0.384,\end{aligned}$$

$$\begin{aligned}
\hat{p}_2(\{1\}) &= 0.227, \hat{p}_2(\{2\}) = 0.182, \hat{p}_2(\{3\}) = 0.394, \\
\hat{p}_2(\{1, 2\}) &= 0.197, \\
\hat{\mu}(\{1\}) &= 1.002, \hat{\mu}(\{2\}) = 1.000, \hat{\mu}(\{3\}) = 1.004, \\
\hat{\sigma}(1, \{1\}) &= 0.950, \hat{\sigma}(1, \{2\}) = 1.393, \hat{\sigma}(1, \{3\}) = 0.999, \\
\hat{\sigma}(2, \{1\}) &= 18.22, \hat{\sigma}(2, \{2\}) = 19.18, \hat{\sigma}(2, \{3\}) = 24.83, \\
\hat{\xi}(1, \{1\}) &= 0.552, \hat{\xi}(1, \{2\}) = 0.507, \hat{\xi}(1, \{3\}) = 0.493, \\
\hat{\xi}(2, \{1\}) &= 2.206, \hat{\xi}(2, \{2\}) = 2.220, \hat{\xi}(2, \{3\}) = 1.888, \\
\hat{\mu}(\{1, 2\}) &= \begin{pmatrix} 3.142 \\ 3.040 \end{pmatrix}, \hat{\sigma}(2, \{1, 2\}) = \begin{pmatrix} 25.98 \\ 18.06 \end{pmatrix}, \hat{\alpha}(2, \{1, 2\}) = 1.79.
\end{aligned}$$

The estimation procedure works quite well and the states are correctly retrieved, see Figure 9. Of course, if the observation period was shorter, or if the phase change intensities were smaller, then it would be impossible to estimate transition rates accurately.

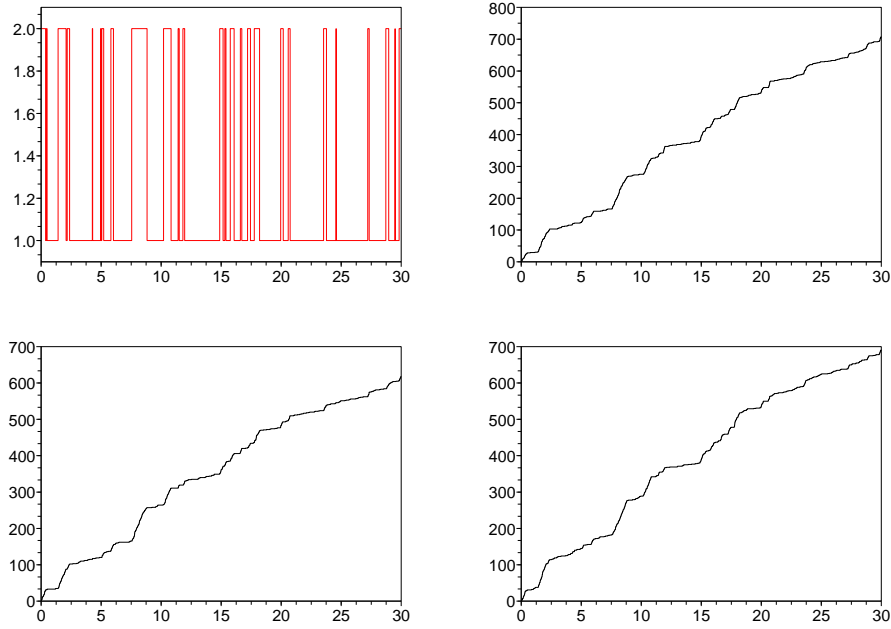


Figure 7: The counting processes: top left: the true process  $J$ , top right: the counting process related to  $S_1$ , bottom left: the counting process related to  $S_2$ , bottom right: the counting process related to  $S_3$

#### 6.4.3 A model with 3 states of the environment

We now assume that  $r = 3$  and that common shocks are not present (for  $i = 1, 2, 3, p_i(e) = 0$  if  $\text{Card}(e) \geq 2$ ). In state 1, claims are not very frequent and not very severe in the three zones. In state 2, claims are more likely and more severe in average for the three zones. State 3 corresponds to exceptional conditions that favor extremely severe claims for zones A

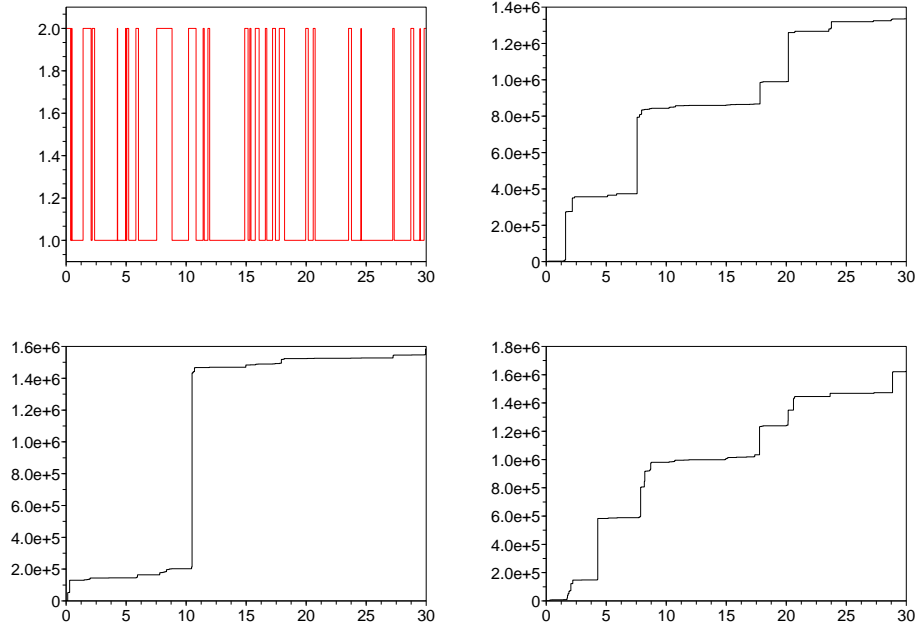


Figure 8: The loss processes  $S_k$ , top left: the true process  $J$ , top right: the loss process  $S_1$ , bottom left: the loss process  $S_2$ , bottom right: the loss process  $S_3$

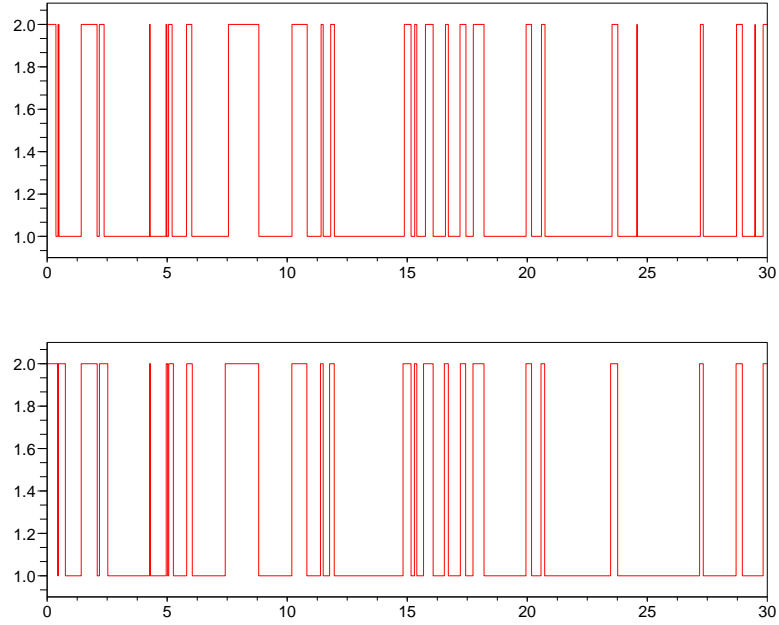


Figure 9: Reconstruction of the hidden Markov process  $J$ : top: the true process  $J$ , bottom: the reconstructed process  $\hat{J}$

and B but protect zone C. Take  $\lambda_1 = 20$ ,  $\lambda_2 = 200$ ,  $\lambda_3 = 1000$ ,  $p_1(\{1\}) = p_1(\{2\}) = 0.3$ ,  $p_1(\{3\}) = 0.4$ ,  $p_2(\{1\}) = p_2(\{2\}) = 0.3$ ,  $p_2(\{3\}) = 0.4$ ,  $p_3(\{1\}) = p_1(\{2\}) = 0.45$  and  $p_1(\{3\}) = 0.1$ . The claim severity distributions are once again modeled by GP distributions, with

$$\begin{aligned}\mu(\{1\}) &= \mu(\{2\}) = \mu(\{3\}) = 1, \\ \sigma(1, \{1\}) &= \sigma(1, \{2\}) = \sigma(1, \{3\}) = 1, \\ \sigma(2, \{1\}) &= \sigma(2, \{2\}) = \sigma(2, \{3\}) = 20, \\ \sigma(3, \{1\}) &= \sigma(3, \{2\}) = 200, \sigma(3, \{3\}) = 0.5, \\ \xi(1, \{1\}) &= \xi(1, \{2\}) = \xi(1, \{3\}) = 1/4, \\ \xi(2, \{1\}) &= \xi(2, \{2\}) = \xi(2, \{3\}) = 1/2, \\ \xi(2, \{1\}) &= \xi(2, \{2\}) = 1, \xi(2, \{3\}) = 1/3,\end{aligned}$$

These parameters are chosen so that claims for zone C in state 3 are very small compared to those for zones A and B. Assume that we observe the multivariate claim process during 30 years, that the average time spent in state 1 (before switching to another state) is 1 year (resp. 3 months for state 2, 1 month for state 3), and that jumps from state 1 to state 3 or from state 3 to state 1 are a.s. impossible. Assume finally that when one leaves state 2, the probability to go to state 1 is  $2/3$ . The intensity transition parameters are then  $\ell_{12} = 1$ ,  $\ell_{13} = 0$ ,  $\ell_{21} = 8/3$ ,  $\ell_{23} = 4/3$ ,  $\ell_{31} = 0$ ,  $\ell_{32} = 12$ .

Again, the estimate of  $\mu(\{i\})$ ,  $i = 1, 2, 3$  is chosen as the minimum of the claims affecting line  $i$ . The results are the following:

$$\begin{aligned}\widehat{\ell}_{12} &= 1.691, \widehat{\ell}_{13} = 0, \widehat{\ell}_{21} = 2.513, \widehat{\ell}_{23} = 1.288, \widehat{\ell}_{31} = 0, \widehat{\ell}_{32} = 10.76, \\ \widehat{\lambda}_1 &= 27.44, \widehat{\lambda}_2 = 198.3, \widehat{\lambda}_3 = 976.3, \\ \widehat{p}_1(\{1\}) &= 0.289, \widehat{p}_1(\{2\}) = 0.332, \widehat{p}_1(\{3\}) = 0.379, \\ \widehat{p}_2(\{1\}) &= 0.306, \widehat{p}_2(\{2\}) = 0.298, \widehat{p}_2(\{3\}) = 0.396, \\ \widehat{p}_3(\{1\}) &= 0.448, \widehat{p}_3(\{2\}) = 0.444, \widehat{p}_3(\{3\}) = 0.109, \\ \widehat{\mu}(\{1\}) &= 1.003, \widehat{\mu}(\{2\}) = 1.001, \widehat{\mu}(\{3\}) = 1.000, \\ \widehat{\sigma}(1, \{1\}) &= 1.013, \widehat{\sigma}(1, \{2\}) = 1.065, \widehat{\sigma}(1, \{3\}) = 1.016, \\ \widehat{\sigma}(2, \{1\}) &= 19.17, \widehat{\sigma}(2, \{2\}) = 19.85, \widehat{\sigma}(2, \{3\}) = 20.83, \\ \widehat{\sigma}(3, \{1\}) &= 191.9, \widehat{\sigma}(3, \{2\}) = 191.2, \widehat{\sigma}(3, \{3\}) = 0.472, \\ \widehat{\xi}(1, \{1\}) &= 0.356, \widehat{\xi}(1, \{2\}) = 0.298, \widehat{\xi}(1, \{3\}) = 0.251, \\ \widehat{\xi}(2, \{1\}) &= 0.504, \widehat{\xi}(2, \{2\}) = 0.437, \widehat{\xi}(2, \{3\}) = 0.433, \\ \widehat{\xi}(2, \{1\}) &= 0.957, \widehat{\xi}(2, \{2\}) = 0.948, \widehat{\xi}(2, \{3\}) = 0.443.\end{aligned}$$

Once again, results are correct because we have enough environment process changes during our observation period, see Figure 12. Results are slightly less accurate than in the 2-dimensional case, for example regarding  $\lambda_1$ . Note that even if results would be completely

inaccurate for large numbers of lines or numbers of states of the environment, estimation and reconstruction results are acceptable for 3 lines and 3 states of the environment.

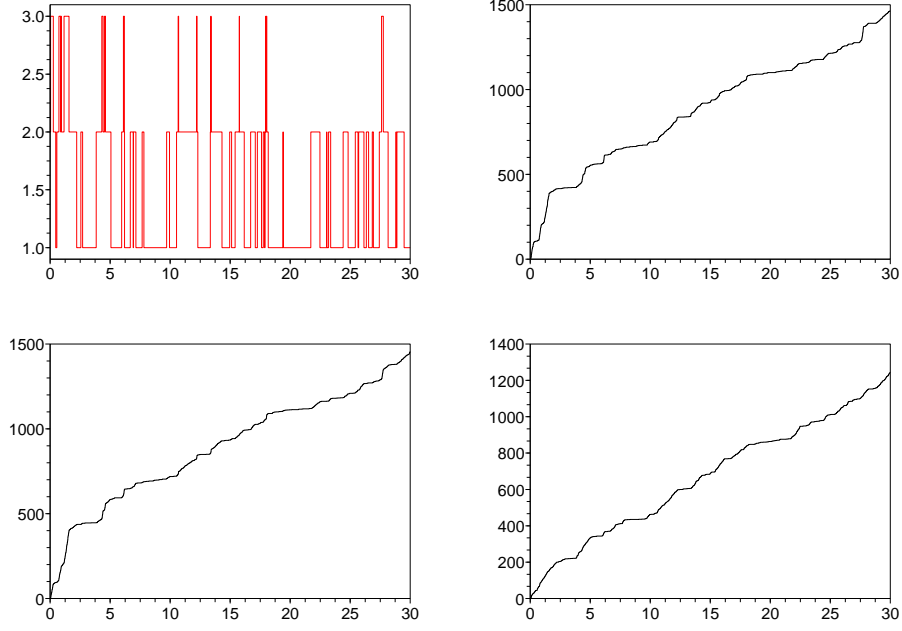


Figure 10: The counting processes: top left: the true process  $J$ , top right: the counting process related to  $S_1$ , bottom left: the counting process related to  $S_2$ , bottom right: the counting process related to  $S_3$

## Acknowledgments

The authors thank very much Alexandre You for his valuable help on this project, and Jean-Baptiste Gouere for a useful comment. The second author acknowledges partial support from the research chair Actuariat Durable sponsored by Milliman, from the research chair Actuariat Responsable sponsored by Generali, and from French Research National Agency (ANR) under the reference ANR-08-BLAN-0314-01.

## References

- [1] Asmussen, S. (1989). Risk theory in a Markovian environment. *Scand. Actuar. J.* **2**, 69–100.
- [2] Asmussen, S. (2000). *Ruin probabilities*, World Scientific.
- [3] Baum, L.E., Petrie, T. (1966). Statistical inference for probabilistic functions of finite-state Markov chains. *Ann. Math. Statist.* **37**, 1554–1563.



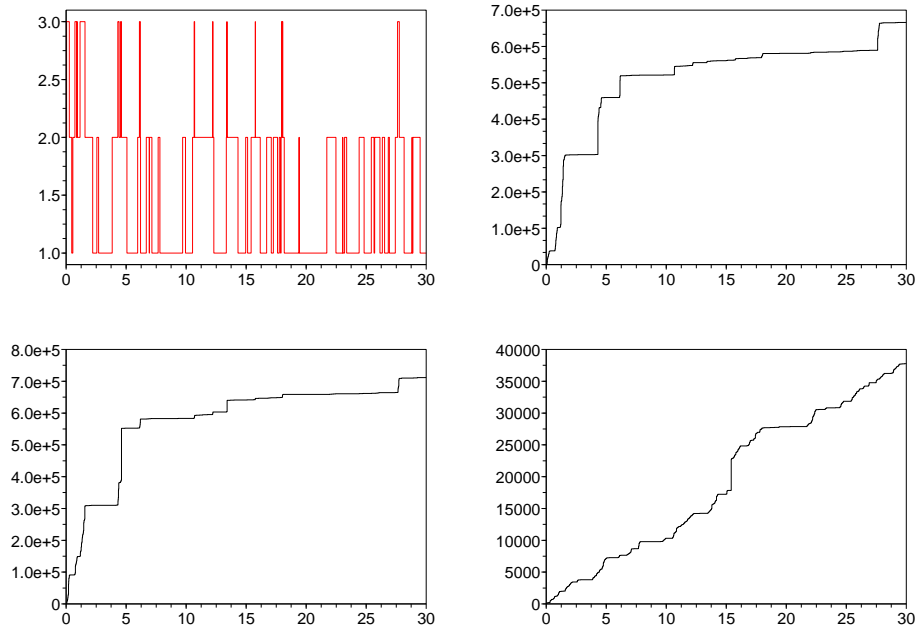


Figure 11: The loss processes  $S_k$ , top left: the true process  $J$ , top right: the loss process  $S_1$ , bottom left: the loss process  $S_2$ , bottom right: the loss process  $S_3$

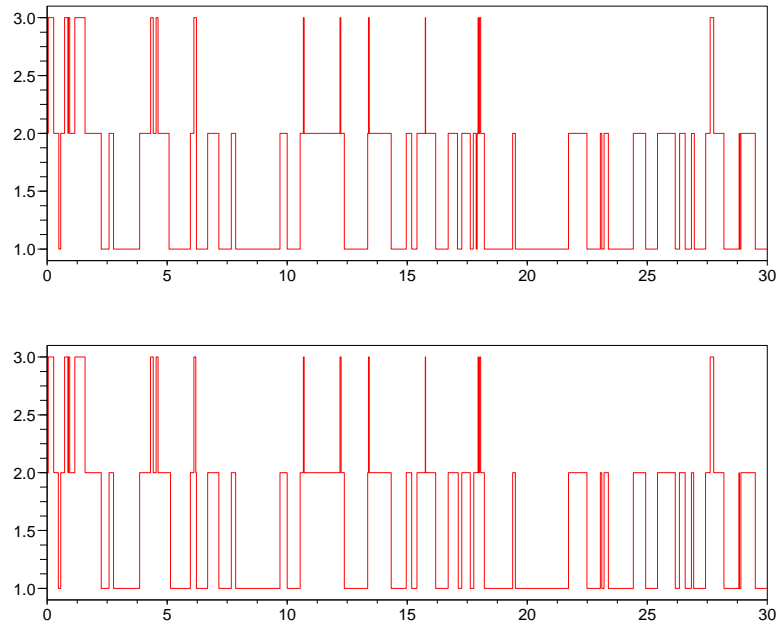


Figure 12: Reconstruction of the hidden Markov process  $J$ : top: the true process  $J$ , bottom: the reconstructed process  $\hat{J}$

- [4] Baum, L.E., Petrie, T., Soules, G., Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.* **41**, 164–171.
- [5] Bickel, P.J., Ritov, Y., Rydén, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Statist.* **26**(4), 1614–1635.
- [6] Buch-Larsen, T., Nielsen, J.P., Guillen, M., Bolanc, C. (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics* **39**(6), 503–516.
- [7] Cai, J., Tan, K.S. (2007). Optimal retention for a stop-loss reinsurance under the VaR and CTE risk measure. *ASTIN Bull.* **37**(1), 93–112.
- [8] Caliebe, A. (2006). Properties of the maximum a posteriori path estimator in hidden Markov models. *IEEE Trans. Inform. Theory* **52**(1), 41–51.
- [9] Champernowne, D.G. (1936). The Oxford Meeting, September 25–29. *Econometrica* **5**, October 1937.
- [10] Champernowne, D.G. (1937). The theory of income distribution, *Econometrica* **5**, 379–381.
- [11] Chiragiev, A., Landsman, Z. (2007). Multivariate Pareto portfolios: TCE-based capital allocation and divided differences. *Scand. Actuar. J.* **2007**(4), 261–280.
- [12] Çinlar, E. (1975). *Introduction to stochastic processes*, Prentice-Hall.
- [13] Davison, A.C., Ramesh, N.I. (1993). A stochastic model for times of exposures to air pollution from a point source, in *Statistics for the environment*, editors: V. Barnett and K.F. Turkman, Wiley, New York.
- [14] Deng, L., Mark, J.W. (1993). Parameter estimation for Markov modulated Poisson processes via the EM algorithm with time discretization. *Telecomm. Syst.* **1**, 321–338.
- [15] Fischer, W., Meier-Hellstern, K.S. (1993). The Markov-modulated Poisson process (MMPP) cookbook. *Perf. Eval.* **18**, 149–171.
- [16] Guillou, A., Loisel, S., Stupfler, G. (2011). Asymptotic normality of the maximum likelihood estimator of a loss process, available on the webpage <http://www-irma.u-strasbg.fr/~guillou/supplement.pdf>.
- [17] Gusella, R. (1991). Characterizing the variability of arrival processes with indexes of dispersion. *IEEE J. Select. Areas Commun.* **9**, 203–211.

- [18] Heffes, H., Lucantoni, D. (1986). A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE J. Select. Areas Commun.* **4**, 856–867.
- [19] Holst, U., Lindgren, G. (1995). Recursive estimation of parameters in Markov-modulated Poisson processes. *IEEE Trans. Commun.* **43**(11), 2812–2820.
- [20] Johnson, N.L., Kotz, S., Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. 1, 2nd. ed., New-York: John Wiley and Sons.
- [21] Kawashima, K., Saito, H. (1990). Teletraffic issues in ATM networks. *Comput. Networks and ZSDN Systems* **20**, 369–375.
- [22] Kingman, J.F.C. (1976). Subadditive processes, in *Ecole d’Eté de Probabilités de Saint-Flour V-1975*, volume 539 of *Lecture Notes in Math.*, Springer-Verlag.
- [23] Leroux, B.G. (1992). Maximum-likelihood estimation for hidden Markov models. *Stoch. Proc. Appl.* **40**, 127–143.
- [24] Lindskog, F., McNeil, A. (2003). Common Poisson shock models: applications to insurance and credit risk modelling. *ASTIN Bull.* **33**(2), 209–238.
- [25] Loisel, S. (2004). Ruin theory with  $K$  lines of business. Proceedings of the 3rd Actuarial and Financial Day, Brussels.
- [26] Loisel, S., Milhaud, X. (2011). From deterministic to stochastic surrender risk models: impact of correlation crises on economic capital, available at <http://hal.archives-ouvertes.fr/hal-00502847>.
- [27] McLachlan, G.J., Krishnan, T. (1997). *The EM Algorithm and Extensions*, Wiley, New York.
- [28] Meier-Hellstern, K.S. (1987). A fitting algorithm for Markov-modulated Poisson processes having two arrival rates. *European J. Oper. Res.* **29**, 970–977.
- [29] Milhaud, X., Loisel, S., Maume-Deschamps, V. (2010). Surrender triggers in life insurance: classification and risk predictions, available at <http://hal.archives-ouvertes.fr/hal-00450003>.
- [30] Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. (1989). *Numerical Recipes*, Cambridge University Press.
- [31] Rossiter, M.H. (1989). *Characterizing a random point process by a switched Poisson process*, Ph.D. thesis, Monash University, Melbourne.

- [32] Rydén, T. (1994a). Parameter estimation for Markov modulated Poisson processes. *Commun. Stat., Stochastic Models* **10**(4), 795–829.
- [33] Rydén, T. (1994b). Consistent and asymptotically normal parameter estimates for hidden Markov models. *Ann. Statist.* **22**(4), 1884–1895.
- [34] Rydén, T. (1995). Consistent and asymptotically normal parameter estimates for Markov modulated Poisson processes. *Scand. J. Stat.* **22**(3), 295–303.
- [35] Rydén, T. (1996). An EM algorithm for estimation in Markov-modulated Poisson processes. *Comput. Statist. Data Anal.* **21**(4), 431–447.
- [36] Vandekerkhove, P. (2005). Consistent and asymptotically normal parameter estimates for hidden Markov mixtures of Markov models. *Bernoulli* **11**(11), 103–129.
- [37] Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inform. Theory* **IT-13**(2), 260–269.

## 7 Appendix A: Auxiliary results

We first state three technical lemmas in order to adapt the intermediate results in [32]:

**Lemma 1.** *Let  $(X, \mathcal{M}, \mu)$  be a finite measure space. Assume that  $f$  and  $g$  are two Borel nonnegative functions on  $X$  such that  $\ln f$  and  $\ln g$  are  $\mu$ -integrable. Then  $\ln(f + g)$  is  $\mu$ -integrable.*

**Lemma 2.** *For all  $\Phi$  and  $\Phi' \in F_\Phi$ , set  $\mathcal{C}(\Phi') = \{i \in \{1, \dots, r\} \mid \pi_i(\Phi') > 0\}$ . Then, for all  $i \in \mathcal{C}(\Phi')$ , one has*

$$\mathbb{E}_\Phi \left[ \max_{i \in \mathcal{C}(\Phi')} |\ln g_i(Y_1, E_1, X_1, \Phi')| \right] < \infty.$$

Recall that a double-index process  $(W_{st})$  is said to be *subadditive* (see [22]) if

1.  $\forall s < u < t \quad W_{st} \leq W_{su} + W_{ut};$
2.  $W_{st}$  is stationary relative to the shift  $W_{st} \rightarrow W_{s+1, t+1};$
3.  $\mathbb{E}|W_{0t}| < \infty;$
4.  $\exists A \in \mathbb{R} \quad \mathbb{E}(W_{0t}) \geq -At.$

In such a case (see [22, Theorems 1.5 and 1.7]), there exists an integrable random variable  $\xi$  such that

1.  $\frac{W_{0t}}{t} \rightarrow \xi$  a.s. and in  $L^1;$

2. If the invariant  $\sigma$ -algebra relative to the shift in 2. is a.s. trivial, then  $\xi$  is a.s. constant.

**Lemma 3.** *Pick an arbitrary parameter  $\Phi$ , and set for all  $\Phi' \in F_\Phi$*

$$q_{st}(\Phi') = \max_{j \in \mathcal{C}(\Phi')} \mathcal{L}((y_i, e_i, x_i)_{s+1 \leq i \leq t}, \Phi' | J(0) = j).$$

*Then the process  $(W_{st} = \ln q_{st}(\Phi'))$  is  $\mathbb{P}_\Phi$ -subadditive, and there exists a neighborhood  $G$  of  $\Phi'$  in  $F_\Phi$  such that for all  $G_{\Phi'} \subset G$  containing  $\Phi'$ , the process  $(Z_{st} = \ln \sup_{\varphi \in G_{\Phi'}} q_{st}(\varphi))$  is  $\mathbb{P}_\Phi$ -subadditive.*

We now write results analogous to the lemmas of [32], in order to show our main theorem.

**Lemma 4.** *For all  $\Phi$ ,  $\ln \mathcal{L}(Y_1, E_1, X_1, \Phi | (Y_i, E_i, X_i)_{i \leq 0})$  is  $\mathbb{P}_\Phi$ -integrable. Let*

$$H(\Phi) = \mathbb{E}_\Phi(-\ln \mathcal{L}(Y_1, E_1, X_1, \Phi | (Y_i, E_i, X_i)_{i \leq 0})).$$

*Then*

1.  $\frac{1}{n} \mathbb{E}_\Phi(\ln \mathcal{L}((Y_i, E_i, X_i)_{1 \leq i \leq n}, \Phi)) \xrightarrow{n \rightarrow \infty} -H(\Phi);$
2.  $\frac{1}{n} \ln \mathcal{L}((Y_i, E_i, X_i)_{1 \leq i \leq n}, \Phi) \xrightarrow{n \rightarrow \infty} -H(\Phi) \quad \mathbb{P}_\Phi\text{-a.s.}$

**Lemma 5.** *For all  $\Phi$  and all  $\Phi' \in F_\Phi$ , there exists a finite real number  $H(\Phi, \Phi')$  such that*

1.  $\frac{1}{n} \mathbb{E}_\Phi(\ln \mathcal{L}((Y_i, E_i, X_i)_{1 \leq i \leq n}, \Phi')) \xrightarrow{n \rightarrow \infty} H(\Phi, \Phi');$
2.  $\frac{1}{n} \mathbb{E}_\Phi \left[ \max_{j \in \mathcal{C}(\Phi')} \mathcal{L}((Y_i, E_i, X_i)_{1 \leq i \leq n}, \Phi' | J(0) = j) \right] \xrightarrow{n \rightarrow \infty} H(\Phi, \Phi');$
3.  $\frac{1}{n} \ln \mathcal{L}((Y_i, E_i, X_i)_{1 \leq i \leq n}, \Phi') \xrightarrow{n \rightarrow \infty} H(\Phi, \Phi') \quad \mathbb{P}_\Phi\text{-a.s.}$
4.  $\frac{1}{n} \max_{j \in \mathcal{C}(\Phi')} \mathcal{L}((Y_i, E_i, X_i)_{1 \leq i \leq n}, \Phi' | J(0) = j) \xrightarrow{n \rightarrow \infty} H(\Phi, \Phi') \quad \mathbb{P}_\Phi\text{-a.s.}$

To adapt Lemma 6 of [32], define

$$\Omega = \{y_i, e_i, x_i, u_1(i-1), \dots, u_r(i-1)\}_{i=1}^\infty$$

where  $y_i \in \mathbb{R}_+$ ,  $e_i \in \{0, 1\}^n$ ,  $x_i \in \mathbb{R}_+^n$ ,  $u_p(i-1) \in [0, 1]$ .  $\Omega$  is equipped with its Borel  $\sigma$ -algebra  $\mathcal{B}$  (to define a topology on  $\Omega$ , use a weighted sum of metrics so that  $\Omega$  is complete and separable).

Define by induction

$$u_j(0, \Phi) = \pi_j(\Phi), \quad u_j(k, \Phi) = \frac{\sum_i g_{ij}(y_k, e_k, x_k, \Phi) u_i(k-1, \Phi)}{\sum_i g_i(y_k, e_k, x_k, \Phi) u_i(k-1, \Phi)},$$

so that Bayes' formula yields  $u_j(k, \Phi) = \mathbb{P}_\Phi(J_k = j | (y_i, e_i, x_i)_{1 \leq i \leq k})$ .

Let now  $\mathcal{P}_n$  be the set of all subsets of  $\{1, \dots, n\}$  and define, for all Borel sets  $B \subset \mathbb{R}_+^k \times \mathcal{P}_n^k \times \mathbb{R}_+^{nk} \times [0, 1]^{rk}$ ,

$$\mathbb{P}_{\Phi, \Phi'}(B) = \mathbb{P}_\Phi((y_i, e_i, x_i, u(i-1, \Phi'))_{1 \leq i \leq k} \in B).$$

We let  $\mathcal{S} : \Omega \rightarrow \Omega$ ,  $(y_i, e_i, x_i, u(i-1, \Phi)) \mapsto (y_{i+1}, e_{i+1}, x_{i+1}, u(i, \Phi))$  be the standard shift transformation, and

$$\tilde{\mathbb{P}}_{\Phi, \Phi'}^{(j)} = \frac{1}{j} \sum_{i=0}^{j-1} \mathbb{P}_{\Phi, \Phi'} \circ \mathcal{S}^{-i}.$$

Finally, we let  $Y_i, E_i, X_i$  and  $U(i-1, \Phi')$  be the corresponding coordinate mappings.

**Lemma 6.** *For all  $\Phi$  and  $\Phi' \in F_\Phi$ , there exist an increasing sequence of integers  $(j_k)$  and a probability measure  $\tilde{\mathbb{P}}_{\Phi, \Phi'}$  on  $\Omega$  such that*

1. *For all  $p$ , the law of  $((Y_i, E_i, X_i, U(i-1, \Phi'))_{1 \leq i \leq p})$  under the probability  $\tilde{\mathbb{P}}_{\Phi, \Phi'}^{(j_k)}$  converges weakly to the law of  $((Y_i, E_i, X_i, U(i-1, \Phi'))_{1 \leq i \leq p})$  under the probability  $\tilde{\mathbb{P}}_{\Phi, \Phi'}$ ;*
2. *The process  $((Y_i, E_i, X_i, U(i-1, \Phi'))_i)$  is  $\tilde{\mathbb{P}}_{\Phi, \Phi'}$ -stationary;*
3. *The process  $((Y_i, E_i, X_i))_i$  has the same law under  $\tilde{\mathbb{P}}_{\Phi, \Phi'}$  and under  $\mathbb{P}_\Phi$ .*

**Lemma 7.** *Let  $\Phi$  be an arbitrary parameter and  $\Phi' \in F_\Phi$ . Then*

$$H(\Phi, \Phi') = \tilde{\mathbb{E}}_{\Phi, \Phi'} \left( \ln \left( \sum_{i=1}^r U_i(0, \Phi') g_i(Y_1, E_1, X_1, \Phi') \right) \right).$$

For all  $\Phi$  and  $\Phi' \in F_\Phi$ , set now  $K(\Phi, \Phi') = H(\Phi, \Phi) - H(\Phi, \Phi')$ : this real number is well defined and finite. We state a result similar to Lemma 8 of [32]:

**Lemma 8.** *Pick  $\Phi$  and  $\Phi' \in F_\Phi$ . Then  $K(\Phi, \Phi') \geq 0$ , and  $(K(\Phi, \Phi') = 0) \Leftrightarrow (\Phi \sim \Phi')$ .*

We finally write a result that solves the E step of the EM algorithm in Section 4:

**Lemma 9.** *Retain the notations of Section 4, and let*

$$A(e) = \bigcup_{i=1}^r A_i(e) = \{j \in \{1, \dots, k\} \mid e_j = e\}.$$

Then

$$\begin{aligned} \widehat{T}_i &= \int_0^T \frac{\mathbb{P}_\varphi(J(v) = i, N(u), S(u), 0 \leq u < v)}{\mathbb{P}_\varphi(N(u), S(u), 0 \leq u \leq T)} \\ &\quad \times \mathbb{P}_\varphi(N(u), S(u), v \leq u \leq T \mid J(v) = i) dv, \\ \widehat{n}_i &= \sum_{q=1}^k \frac{\mathbb{P}_\varphi(J(t_q) = i, N(u), S(u), 0 \leq u \leq T)}{\mathbb{P}_\varphi(N(u), S(u), 0 \leq u \leq T)}, \\ \widehat{\mathbb{1}_{\{j \in A_i(e)\}}} &= \mathbb{1}_{\{j \in A(e)\}} \mathbb{P}_\varphi(J(t_j) = i \mid N(u), S(u), 0 \leq u \leq T), \\ \widehat{\text{card}(A_i(e))} &= \sum_{j=1}^k \mathbb{1}_{\{j \in A(e)\}} \mathbb{P}_\varphi(J(t_j) = i \mid N(u), S(u), 0 \leq u \leq T), \\ \widehat{m_{ij}(T)} &= \ell_{ij}(\varphi) \int_0^T \frac{\mathbb{P}_\varphi(J(v) = i, N(u), S(u), 0 \leq u < v)}{\mathbb{P}_\varphi(N(u), S(u), 0 \leq u \leq T)} \\ &\quad \times \mathbb{P}_\varphi(N(u), S(u), v \leq u \leq T \mid J(v) = j) dv. \end{aligned}$$

## 8 Appendix B: Proofs

**Proof of Lemma 1.** Start by writing

$$\int_X |\ln(f+g)| d\mu = \int_{\{f+g \geq 1\}} \ln(f+g) d\mu - \int_{\{f+g < 1\}} \ln(f+g) d\mu.$$

Now  $f \leq f+g$ , so that  $-\ln(f+g)\mathbb{1}_{\{f+g < 1\}} \leq -\ln f \mathbb{1}_{\{f+g < 1\}}$ . Therefore

$$\int_X |\ln(f+g)| d\mu \leq \int_{\{f+g \geq 1\}} \ln(f+g) d\mu + \int_X |\ln f| d\mu.$$

Thus, it is enough to prove that  $\int_{\{f+g \geq 1\}} \ln(f+g) d\mu < \infty$ .

Notice that if  $x, y$  are positive real numbers  $\geq 2$ , then

$$\ln\left(\frac{1}{x} + \frac{1}{y}\right) \leq 0$$

which leads to

$$\ln(x+y) \leq \ln x + \ln y.$$

Consequently

$$\begin{aligned} \int_{\{f+g \geq 1\}} \ln(f+g) d\mu &= \int_{\{f+g \geq 1, f \geq 2, g \geq 2\}} \ln(f+g) d\mu + \int_{\{f+g \geq 1, f < 2, g < 2\}} \ln(f+g) d\mu \\ &\quad + \int_{\{f+g \geq 1, f \geq 2, g < 2\}} \ln(f+g) d\mu + \int_{\{f+g \geq 1, f < 2, g \geq 2\}} \ln(f+g) d\mu \\ &\leq \int_{\{f+g \geq 1, f \geq 2, g \geq 2\}} (\ln f + \ln g) d\mu + \mu(X) \cdot 2 \ln 2 \\ &\quad + \int_{\{f+g \geq 1, f \geq 2, g < 2\}} \ln(f+2) d\mu + \int_{\{f+g \geq 1, f < 2, g \geq 2\}} \ln(g+2) d\mu \\ &\leq 2 \int_X [|\ln f| + |\ln g|] d\mu + \mu(X) \cdot 4 \ln 2 \end{aligned}$$

which gives  $\int_{\{f+g \geq 1\}} \ln(f+g) d\mu < \infty$  and ends the proof of this lemma. ■

**Proof of Lemma 2.** First, note that

$$\mathbb{E}_\Phi \left[ \max_{i \in \mathcal{C}(\Phi')} |\ln g_i(Y_1, E_1, X_1, \Phi')| \right] \leq \sum_{i \in \mathcal{C}(\Phi')} \mathbb{E}_\Phi |\ln g_i(Y_1, E_1, X_1, \Phi')|;$$

we shall then prove that for all  $\Phi' \in F_\Phi$  and  $i \in \mathcal{C}(\Phi')$ ,  $\mathbb{E}_\Phi |\ln g_i(Y_1, E_1, X_1, \Phi')| < \infty$ . Before proceeding, notice that, denoting by  $K = \max_{i, e} \lambda_i(\Phi') p_i(e, \Phi') > 0$ , one has

$$g_{ij}(y, e, x, \Phi') \leq K \cdot \mathbb{P}_{\theta(j, e, \Phi')}(m \in e \Rightarrow X_m = x_m).$$

Now, for all  $\Phi$  and  $\Phi' \in F_\Phi$ ,

$$\mathbb{E}_\Phi |\ln g_i(Y_1, E_1, X_1, \Phi')| = \sum_e \int |\ln g_i(y, e, x, \Phi')| \mathcal{L}(dy, e, dx, \Phi).$$

Using the equality  $g_i = \sum_j g_{ij}$  and Lemma 1, it is enough to show that for all  $i$  and all  $j, e$  such that  $\lambda_j(\Phi')p_j(e, \Phi') > 0$ , the function

$$(y, x) \mapsto \ln \left[ \exp(y(L(\Phi') - \Lambda(\Phi')))_{i,j} \lambda_j(\Phi')p_j(e, \Phi') \mathbb{P}_{\theta(j,e,\Phi')}(m \in e \Rightarrow X_m = x_m) \right]$$

is  $\mathcal{L}(\cdot, e, \cdot, \Phi)$ -integrable. The hypothesis on the statistical model  $(\mathbb{P}_\theta)$  first gives, for all  $i, j$ ,

$$\int |\ln \mathbb{P}_{\theta(j,e,\Phi')}(m \in e \Rightarrow X_m = x_m)| \mathbb{P}_{\theta(i,e,\Phi)}(m \in e \Rightarrow X_m \in dx_m) < \infty.$$

Because  $J$  is a  $\mathbb{P}_{\Phi'}$ -irreducible Markov process, for all  $j \neq i$ , there exists an integer  $n(i, j)$  such  $(L(\Phi') - \Lambda(\Phi'))_{i,j}^{n(i,j)} > 0$ , which we pick minimal among the integers satisfying this property. Consequently, in a neighborhood of 0,

$$\exp(y(L(\Phi') - \Lambda(\Phi')))_{i,j} = (L(\Phi') - \Lambda(\Phi'))_{i,j}^{n(i,j)} \cdot \frac{y^{n(i,j)}}{n(i,j)!} + O(y^{n(i,j)+1})$$

so that the functions

$$y \mapsto \ln \left[ \exp(y(L(\Phi') - \Lambda(\Phi')))_{i,j} \lambda_j(\Phi')p_j(e, \Phi') \right] \mathcal{L}(y, e, \Phi)$$

are integrable in a neighborhood of 0.

Using a corollary of Perron-Frobenius' theorem (see e.g. [2, Appendix A.4.8]) entails that there exists a matrix  $A$  with positive entries, and two real numbers  $\mu > \delta' > 0$  such that, as  $y \rightarrow +\infty$ ,

$$\exp(y(L(\Phi') - \Lambda(\Phi'))) = \exp(-\delta'y)A + O(\exp(-\mu y));$$

thus, for all  $i, j \in \{1, \dots, r\}$ , there exists  $a_{i,j} > 0$  such that

$$\exp(y(L(\Phi') - \Lambda(\Phi')))_{i,j} = \exp(-\delta'y)a_{i,j} + O(\exp(-\mu y)).$$

Therefore, there exists  $\delta, c > 0$  such that  $\exp(y(L(\Phi') - \Lambda(\Phi')))_{i,j} \leq c \exp(-\delta y)$ , and the functions

$$y \mapsto \ln \left[ \exp(y(L(\Phi') - \Lambda(\Phi')))_{i,j} \lambda_j(\Phi')p_j(e, \Phi') \right] \mathcal{L}(y, e, \Phi)$$

are integrable in a neighborhood of  $+\infty$ . Finally, Lemma 2 follows.

**Proof of Lemma 3.** Focus first on the process  $W$ . A proof similar to the one of [32] shows that the first and second hypothesis of Kingman's theorem hold. To prove the third and fourth ones, write

$$q_{0p}(\Phi') = \max_{k_0 \in \mathcal{C}(\Phi')} \sum_{k_1, \dots, k_p \in \mathcal{C}(\Phi')} \prod_{s=1}^p g_{k_{s-1}, k_s}(Y_s, E_s, X_s, \Phi'),$$



so that, because  $\left| \ln \max_{1 \leq i \leq n} a_i \right| \leq \max_{1 \leq i \leq n} |\ln a_i|$ ,

$$\begin{aligned} |\ln q_{0p}(\Phi')| &\leq \max_{k_0 \in \mathcal{C}(\Phi')} \left| \ln \left[ \sum_{k_1, \dots, k_p \in \mathcal{C}(\Phi')} \prod_{s=1}^p g_{k_{s-1}, k_s}(Y_s, E_s, X_s, \Phi') \right] \right| \\ &\leq \sum_{k_0 \in \mathcal{C}(\Phi')} \left| \ln \left[ \sum_{k_1, \dots, k_p \in \mathcal{C}(\Phi')} \prod_{s=1}^p g_{k_{s-1}, k_s}(Y_s, E_s, X_s, \Phi') \right] \right|. \end{aligned}$$

To show that the third hypothesis holds, that is  $\mathbb{E}_\Phi |\ln q_{0p}(\Phi')| < \infty$ , we notice that applying Lemma 1 and using the properties of the logarithm function, it is enough to show that for all  $s, e_s$  and  $k_{s-1}, k_s \in \mathcal{C}(\Phi')$  such that  $\lambda_{k_s}(\Phi') p_{k_s}(e_s, \Phi') > 0$ ,

$$\int |\ln g_{k_{s-1}, k_s}(y_s, e_s, x_s, \Phi')| \mathcal{L}((dy_j, e_j, dx_j)_{1 \leq j \leq p}, \Phi) < \infty,$$

which is exactly Lemma 2, and thus the third requirement is met.

Let us now justify that the fourth hypothesis holds. Using the definition of  $F_\Phi$  and the convergence  $\lim_{x \rightarrow 0^+} x \ln x = 0$ , we see that we may restrict the sums to the indexes  $e_j$  such that there exists an index  $i \in \{1, \dots, r\}$  with  $\lambda_i(\Phi') p_i(e_j, \Phi') > 0$ . Let  $I$  be the set of these indexes.

Put then  $\gamma = \min\{\lambda_j(\Phi') p_j(e, \Phi'), j, e \text{ such that } \lambda_j(\Phi') p_j(e, \Phi') > 0\}$ . We set indexes  $e_j \in I$  for all  $j \in \{1, \dots, p\}$ . There exist  $k_1(e_1), \dots, k_p(e_p)$  such that

$$\forall j \in \{1, \dots, p\} \quad \lambda_{k_j(e_j)}(\Phi') p_{k_j(e_j)}(e_j, \Phi') \geq \gamma > 0.$$

Since

$$\begin{aligned} q_{0p}(\Phi') &= \max_{j_0 \in \mathcal{C}(\Phi')} \sum_{j_1, \dots, j_p \in \mathcal{C}(\Phi')} \prod_{s=1}^p g_{j_{s-1}, j_s}(Y_s, E_s, X_s, \Phi') \\ &\geq \max_{k_0 \in \mathcal{C}(\Phi')} \prod_{s=1}^p g_{k_{s-1}(E_{s-1}), k_s(E_s)}(Y_s, E_s, X_s, \Phi') \end{aligned}$$

and

$$\begin{aligned} &g_{k_{s-1}(e_{s-1}), k_s(e_s)}(y_s, e_s, x_s, \Phi') \\ &\geq \gamma \exp(y_s(L(\Phi') - \Lambda(\Phi')))_{k_{s-1}(e_{s-1}), k_s(e_s)} \mathbb{P}_{\theta(k_s(e_s), e_s, \Phi')}(m \in e_s \Rightarrow X_m = x_{m,s}) \end{aligned}$$

we get, for all  $k_0 = k_0(e_0) \in \mathcal{C}(\Phi')$ ,

$$\begin{aligned} \mathbb{E}_\Phi \ln q_{0p}(\Phi') &\geq pC \ln \gamma + \sum_{\substack{e_s \in I \\ 1 \leq s \leq p}} \int \ln \exp(y_s(L(\Phi') - \Lambda(\Phi')))_{k_{s-1}(e_{s-1}), k_s(e_s)} \mathcal{L}((dy_j, e_j, dx_j)_{1 \leq j \leq p}, \Phi) \\ &\quad + \sum_{\substack{e_s \in I \\ 1 \leq s \leq p}} \int \ln \mathbb{P}_{\theta(k_s(e_s), e_s, \Phi')}(m \in e_s \Rightarrow X_m = x_{m,s}) \mathcal{L}((dy_j, e_j, dx_j)_{1 \leq j \leq p}, \Phi) \end{aligned}$$

where  $C = \text{card}(I)$ . We can eliminate the condition  $e_j \in I$  because the terms with  $e_j \notin I$  do not contribute, and thus

$$\begin{aligned} \mathbb{E}_\Phi \ln q_{0p}(\Phi') &\geq pC \ln \gamma + \sum_{\substack{e_s \\ 1 \leq s \leq p}} \int \ln \exp(y_s(L(\Phi') - \Lambda(\Phi')))_{k_{s-1}(e_{s-1}), k_s(e_s)} \mathcal{L}(dy_s, e_{s-1}, e_s, \Phi) \\ &\quad + \sum_{\substack{e_s \\ 1 \leq s \leq p}} \int \ln \mathbb{P}_{\theta(k_s(e_s), e_s, \Phi')}(m \in e_s \Rightarrow X_m = x_{m,s}) \mathcal{L}(e_s, dx_s, \Phi). \end{aligned}$$

The stationarity of the process  $((Y_i, E_i, X_i))_i$  yields

$$\begin{aligned} &\left| \sum_{\substack{e_s \\ 1 \leq s \leq p}} \int \ln \mathbb{P}_{\theta(k_s(e_s), e_s, \Phi')}(m \in e_s \Rightarrow X_m = x_{m,s}) \mathcal{L}(e_s, dx_s, \Phi) \right| \\ &\leq p \sum_{e_1} \int |\ln \mathbb{P}_{\theta(k_1(e_1), e_1, \Phi')}(m \in e_1 \Rightarrow X_m = x_{m,1})| \mathcal{L}(e_1, dx_1, \Phi) < \infty \end{aligned}$$

and

$$\begin{aligned} &\left| \sum_{\substack{e_s \\ 2 \leq s \leq p}} \int \ln \exp(y_s(L(\Phi') - \Lambda(\Phi')))_{k_{s-1}(e_{s-1}), k_s(e_s)} \mathcal{L}(dy_s, e_{s-1}, e_s, \Phi) \right| \\ &\leq (p-1) \sum_{e_1, e_2} \int |\ln \exp(y_2(L(\Phi') - \Lambda(\Phi')))_{k_1(e_1), k_2(e_2)}| \mathcal{L}(dy_2, e_1, e_2, \Phi) < \infty. \end{aligned}$$

The term with  $s = 1$  is also finite, which ensures that  $W$  meets the fourth requirement.

We now turn to the process  $Z$ . Once again, a proof similar to the one of [32] shows that the first and second hypothesis are satisfied for any  $G$ . To show the third one, let  $G$  be an arbitrary subset of  $F_\Phi$  containing  $\Phi'$ . Set  $B = \{\sup_{\varphi \in G} q_{0p}(\varphi) \leq 1\}$ . The inequality

$$q_{0p}(\varphi) \leq \prod_{i=1}^p q_{i-1, i}(\varphi)$$

and the stationarity of the process  $((Y_i, E_i, X_i))_i$  together imply

$$\begin{aligned} \mathbb{E}_\Phi \left| \ln \sup_{\varphi \in G} q_{0p}(\varphi) \right| &= -\mathbb{E}_\Phi \left[ \ln \sup_{\varphi \in G} q_{0p}(\varphi) \mathbb{1}_B \right] + \mathbb{E}_\Phi \left[ \ln \sup_{\varphi \in G} q_{0p}(\varphi) \mathbb{1}_{B^c} \right] \\ &\leq \mathbb{E}_\Phi |\ln q_{0p}(\Phi')| + p \mathbb{E}_\Phi \left| \ln \sup_{\varphi \in G} q_{01}(\varphi) \right|. \end{aligned}$$

Put then  $C = \{\sup_{\varphi \in G} q_{01}(\varphi) \leq 1\}$  and write

$$\begin{aligned} \mathbb{E}_\Phi \left| \ln \sup_{\varphi \in G} q_{01}(\varphi) \right| &= -\mathbb{E}_\Phi \left[ \ln \sup_{\varphi \in G} q_{01}(\varphi) \mathbb{1}_C \right] + \mathbb{E}_\Phi \left[ \ln \sup_{\varphi \in G} q_{01}(\varphi) \mathbb{1}_{C^c} \right] \\ &\leq \mathbb{E}_\Phi |\ln q_{01}(\Phi')| + \mathbb{E}_\Phi \left[ \ln \sup_{\varphi \in G} q_{01}(\varphi) \mathbb{1}_{C^c} \right]. \end{aligned}$$

Because  $\forall p \ \mathbb{E}_\Phi |\ln q_{0p}(\Phi')| < \infty$ , it is sufficient to prove that there exists a neighborhood  $G$  of  $\Phi'$  in  $F_\Phi$  such that for every subset  $G_{\Phi'}$  of  $G$  containing  $\Phi'$ ,

$$\mathbb{E}_\Phi \left[ \ln \sup_{\varphi \in G_{\Phi'}} q_{01}(\varphi) \mathbb{1}_{C^c} \right] < \infty.$$

Since

$$q_{01}(\varphi) = \max_{i \in \mathcal{C}(\varphi)} g_i(Y_1, E_1, X_1, \varphi) \leq \sum_{1 \leq i \leq r} g_i(Y_1, E_1, X_1, \varphi)$$

one has

$$\sup_{\varphi \in G} q_{01}(\varphi) \leq \sum_{1 \leq i \leq r} \sum_{1 \leq j \leq r} \sup_{\varphi \in G} g_{ij}(Y_1, E_1, X_1, \varphi).$$

Lemma 1 therefore shows that it is enough to find a neighborhood  $G$  of  $\Phi'$  in  $F_\Phi$  such that for all  $e \in I$ , all  $j$  such that  $\lambda_j(\Phi')p_j(e, \Phi') > 0$  and every subset  $G_{\Phi'}$  of  $G$  containing  $\Phi'$ , we have

$$\int \left| \ln \sup_{\varphi \in G_{\Phi'}} g_{ij}(y, e, x, \varphi) \right| \mathcal{L}(dy, e, dx, \Phi) < \infty$$

for all  $i \in \{1, \dots, r\}$ .

The hypothesis on the statistical model  $(\mathbb{P}_\theta)$  and the fact that  $J$  has a finite state space together imply that there exists a neighborhood  $G_2$  of  $\Phi'$  such that for all  $i, j$  and every  $G_{\Phi'}$  of  $G_2$  containing  $\Phi'$ ,

$$\int \left| \ln \sup_{\varphi \in G_{\Phi'}} \mathbb{P}_{\theta(i, e, \varphi)}(m \in e \Rightarrow X_m = x_m) \right| \mathbb{P}_{\theta(j, e, \Phi)}(m \in e \Rightarrow X_m = x_m) dx < \infty;$$

it is finally enough to find, for every  $e$ , a neighborhood  $G_1$  of  $\Phi'$  such that for every subset  $G_{\Phi'}$  of  $G_1$  containing  $\Phi'$ ,

$$\int \left| \ln \sup_{\varphi \in G_{\Phi'}} \exp(y(L(\varphi) - \Lambda(\varphi)))_{i, j} \cdot \lambda_j(\varphi)p_j(e, \varphi) \right| \mathcal{L}(dy, e, \Phi) < \infty.$$

Pick then a neighborhood  $G_1$  of  $\Phi'$  such that

1.  $\forall i \neq j \quad \sup_{\varphi \in G} |\ell_{ij}(\varphi) - \ell_{ij}(\Phi')| < \ell_{ij}(\Phi') \mathbb{1}_{\{\ell_{ij}(\Phi') > 0\}} + \mathbb{1}_{\{\ell_{ij}(\Phi') = 0\}};$
2.  $\forall j \quad \sup_{\varphi \in G} |\lambda_j(\varphi) - \lambda_j(\Phi')| < \lambda_j(\Phi') \mathbb{1}_{\{\lambda_j(\Phi') > 0\}} + \mathbb{1}_{\{\lambda_j(\Phi') = 0\}};$
3.  $\forall e, j \quad \sup_{\varphi \in G} |p_j(e, \varphi) - p_j(e, \Phi')| < p_j(e, \Phi') \mathbb{1}_{\{p_j(e, \Phi') > 0\}} + \mathbb{1}_{\{p_j(e, \Phi') = 0\}},$

so that if  $\varphi \in G = G_1 \cap G_2$ , the parameters are bounded, and if one of the parameters is positive under  $\mathbb{P}_{\Phi'}$ , then it is also under  $\mathbb{P}_\varphi$ . Set

$$M = \sup_{\varphi \in G} \max_{j, e} \lambda_j(\varphi)p_j(e, \varphi).$$

For all  $\varphi \in G$ , the inequalities

$$\varepsilon \exp(y(L(\varphi) - \Lambda(\varphi)))_{i, j} \leq \exp(y(L(\varphi) - \Lambda(\varphi)))_{i, j} \cdot \lambda_j(\varphi)p_j(e, \varphi) \leq M$$

hold, if  $\varepsilon > 0$  is the infimum of the  $\lambda_k(\varphi)p_k(e, \varphi)$ , for  $\varphi \in G$  and  $e$  such that there exists  $k$  with  $\lambda_k(\Phi)p_k(e, \Phi) > 0$ . Therefore, for all  $\varphi \in G$ ,

$$\varepsilon \exp(y(L(\varphi) - \Lambda(\varphi)))_{i, j} \leq \sup_{\varphi \in G} \exp(y(L(\varphi) - \Lambda(\varphi)))_{i, j} \cdot \lambda_j(\varphi)p_j(e, \varphi) \leq M.$$

Especially,

$$\ln(\varepsilon \exp(y(L(\Phi') - \Lambda(\Phi'))_{i,j})) \leq \ln \left[ \sup_{\varphi \in G} \exp(y(L(\varphi) - \Lambda(\varphi))_{i,j} \cdot \lambda_j(\varphi) p_j(e, \varphi)) \right] \leq \ln M$$

and because the function

$$y \mapsto \ln(\varepsilon \exp(y(L(\Phi') - \Lambda(\Phi'))_{i,j})) \mathcal{L}(y, e, \Phi)$$

is integrable, the result is proved. Finally, noticing that  $Z_{0t} \geq W_{0t}$ , it is clear that the fourth hypothesis holds for  $Z$ , which completes the proof of our result.  $\blacksquare$

**Proof of Lemma 4.** To show Lemma 4, notice that applying Lemma 2 yields

$$\mathbb{E}_{\Phi} \left[ \max_{i \in \mathcal{C}(\Phi)} |\ln g_i(Y_1, E_1, X_1, \Phi)| \right] < \infty$$

and adapt the proof of Lemma 4 in [32].  $\blacksquare$

**Proof of Lemma 5.** Retaining the notation of Lemma 3, the process  $(W_{st})$  is subadditive and ergodic; the proof is then entirely similar to the proof of Lemma 5 in [32] and is therefore omitted.  $\blacksquare$

**Proof of Lemma 6.** The argument goes through as in the proof of Lemma 6 in [32] and is omitted.  $\blacksquare$

**Proof of Lemma 7.** The proof is the same as the proof of Lemma 7 in [32].  $\blacksquare$

**Proof of Lemma 8.** The proof is a straightforward adaptation of the proof of Lemma 8 in [32].  $\blacksquare$

**Proof of Lemma 9.** Since  $\{N(u), S(u), v \leq u \leq T\}$  and  $\{N(u), S(u), 0 \leq u < v\}$  are independent given  $\{J(v) = i\}$ ,

$$\begin{aligned} \hat{T}_i &= \int_0^T \mathbb{P}_{\varphi}(J(v) = i \mid N(u), S(u), 0 \leq u \leq T) dv \\ &= \int_0^T \frac{\mathbb{P}_{\varphi}(J(v) = i, N(u), S(u), 0 \leq u \leq T)}{\mathbb{P}_{\varphi}(N(u), S(u), 0 \leq u \leq T)} dv \\ &= \int_0^T \frac{\mathbb{P}_{\varphi}(J(v) = i, N(u), S(u), 0 \leq u < v)}{\mathbb{P}_{\varphi}(N(u), S(u), 0 \leq u \leq T)} \\ &\quad \times \mathbb{P}_{\varphi}(N(u), S(u), v \leq u \leq T \mid J(v) = i) dv. \end{aligned}$$

To get  $\hat{n}_i$ , write

$$\begin{aligned} \hat{n}_i &= \sum_{q=1}^k \mathbb{P}_{\varphi}(J(t_q) = i \mid N(u), S(u), 0 \leq u \leq T) \\ &= \sum_{q=1}^k \frac{\mathbb{P}_{\varphi}(J(t_q) = i, N(u), S(u), 0 \leq u \leq T)}{\mathbb{P}_{\varphi}(N(u), S(u), 0 \leq u \leq T)}. \end{aligned}$$

Then

$$\widehat{\mathbb{1}_{\{j \in A_i(e)\}}} = \mathbb{1}_{\{j \in A(e)\}} \mathbb{P}_\varphi(J(t_j) = i \mid N(u), S(u), 0 \leq u \leq T).$$

Consequently, we have

$$\widehat{\text{card}(A_i(e))} = \sum_{j=1}^k \widehat{\mathbb{1}_{\{j \in A_i(e)\}}} = \sum_{j=1}^k \mathbb{1}_{\{j \in A(e)\}} \mathbb{P}_\varphi(J(t_j) = i \mid N(u), S(u), 0 \leq u \leq T).$$

We finish by showing how to compute  $\widehat{m_{ij}(T)}$ . First, if  $(U_j)$  stands for the sequence of the jump times of  $J$ ,

$$\begin{aligned} m_{ij}(T) &= \text{card}(\{v: 0 < v \leq T \mid J(v_-) = i, J(v) = j\}) \\ &= \text{card}(\{Q \in \mathbb{N} \setminus \{0\} \mid J(U_{Q-1}) = i, J(U_Q) = j, U_Q \leq T\}) \\ &= \sum_{Q=1}^{+\infty} \mathbb{1}_{\{J(U_{Q-1})=i, J(U_Q)=j, U_Q \leq T\}} \end{aligned}$$

and therefore

$$\begin{aligned} \widehat{m_{ij}(T)} &= \sum_{Q=1}^{+\infty} \mathbb{P}_\varphi(J(U_{Q-1}) = i, J(U_Q) = j, U_Q \leq T \mid N(u), S(u), 0 \leq u \leq T) \\ &= \sum_{Q=1}^{+\infty} \frac{\mathbb{P}_\varphi(J(U_{Q-1}) = i, J(U_Q) = j, U_Q \leq T, N(u), S(u), 0 \leq u \leq T)}{\mathbb{P}_\varphi(N(u), S(u), 0 \leq u \leq T)} \\ &= \int_0^T \sum_{Q=1}^{+\infty} \frac{\mathbb{P}_\varphi(J(U_{Q-1}) = i, J(v) = j, U_Q \in dv, N(u), S(u), 0 \leq u \leq T)}{\mathbb{P}_\varphi(N(u), S(u), 0 \leq u \leq T)} dv. \end{aligned}$$

Now, using independence given  $\{J(v) = j\}$ , one has for all  $Q \in \mathbb{N} \setminus \{0\}$ ,

$$\begin{aligned} \mathbb{P}_\varphi(J(U_{Q-1}) = i, J(v) = j, U_Q \in dv, N(u), S(u), 0 \leq u \leq T) \\ = \mathbb{P}_\varphi(N(u), S(u), v \leq u \leq T \mid J(v) = j) \\ \times \mathbb{P}_\varphi(J(U_{Q-1}) = i, J(v) = j, U_Q \in dv, N(u), S(u), 0 \leq u < v). \end{aligned}$$

Notice that for all  $Q \in \mathbb{N} \setminus \{0\}$ ,

$$\begin{aligned} \mathbb{P}_\varphi(J(U_Q) = j, U_Q - U_{Q-1} \in dv \mid J(U_{Q-1}) = i) &= \frac{\ell_{ij}(\varphi)}{-\ell_{ii}(\varphi)} (-\ell_{ii}(\varphi) \exp(\ell_{ii}(\varphi) v)) dv \\ \mathbb{P}_\varphi(U_Q - U_{Q-1} > v \mid J(U_{Q-1}) = i) &= \exp(\ell_{ii}(\varphi) v) \end{aligned}$$

so that

$$\mathbb{P}_\varphi(J(U_Q) = j, U_Q - U_{Q-1} \in dv \mid J(U_{Q-1}) = i) = \ell_{ij}(\varphi) \mathbb{P}_\varphi(U_Q - U_{Q-1} > v \mid J(U_{Q-1}) = i) dv.$$

Write then

$$\begin{aligned} \mathbb{P}_\varphi(J(U_{Q-1}) = i, J(U_Q) = j, U_Q \in dv, N(u), S(u), 0 \leq u < v) \\ = \int \sum_{j_0, \dots, j_{Q-2}} \mathbb{P}_\varphi \left( \bigcap_{q=0}^{Q-2} \{J(s_q) = j_q\}, \bigcap_{q=0}^{Q-1} \{U_q \in ds_q\}, J(U_{Q-1}) = i, J(U_Q) = j, \right. \end{aligned}$$

$$U_Q \in dv, N(u), S(u), 0 \leq u < v \Bigg).$$

Given  $\left\{ \bigcap_{q=0}^{Q-2} \{J(s_q) = j_q\}, \bigcap_{q=0}^{Q-1} \{U_q \in ds_q\}, J(U_{Q-1}) = i, J(U_Q) = j, U_Q \in dv \right\}$ , the distribution of  $\{N(u), S(u), 0 \leq u < v\}$  is the same as the law of  $\{N(u), S(u), 0 \leq u < v\}$  given  $\left\{ \bigcap_{q=0}^{Q-2} \{J(s_q) = j_q\}, \bigcap_{q=0}^{Q-1} \{U_q \in ds_q\}, J(U_{Q-1}) = i, U_Q > v \right\}$ , because the jump intensities of  $N$  and the shock probabilities in the interval  $[0, v[$  are determined by the values of  $J$  in  $[0, v[$ . Consequently, using Bayes' formula,

$$\begin{aligned} & \mathbb{P}_\varphi(J(U_{Q-1}) = i, J(U_Q) = j, U_Q \in dv, N(u), S(u), 0 \leq u < v) \\ &= \int \sum_{j_0, \dots, j_{Q-2}} \ell_{ij}(\varphi) \cdot \mathbb{P}_\varphi \left( \bigcap_{q=0}^{Q-2} \{J(s_q) = j_q\}, \bigcap_{q=0}^{Q-1} \{U_q \in ds_q\}, J(U_{Q-1}) = i, U_Q > v, \right. \\ & \quad \left. N(u), S(u), 0 \leq u < v \right) dv \end{aligned}$$

which yields

$$\begin{aligned} & \mathbb{P}_\varphi(J(U_{Q-1}) = i, J(U_Q) = j, U_Q \in dv, N(u), S(u), 0 \leq u < v) \\ &= \ell_{ij}(\varphi) \cdot \mathbb{P}_\varphi(J(U_{Q-1}) = i, U_{Q-1} \leq v, U_Q > v, N(u), S(u), 0 \leq u < v) dv, \end{aligned}$$

thus implying that

$$\begin{aligned} & \sum_{Q=1}^{+\infty} \mathbb{P}_\varphi(J(U_{Q-1}) = i, J(U_Q) = j, U_Q \in dv, N(u), S(u), 0 \leq u < v) \\ &= \ell_{ij}(\varphi) \sum_{Q=1}^{+\infty} \mathbb{P}_\varphi(J(U_{Q-1}) = i, U_{Q-1} \leq v, U_Q > v, N(u), S(u), 0 \leq u < v) dv \\ &= \ell_{ij}(\varphi) \mathbb{P}_\varphi(J(v) = i, N(u), S(u), 0 \leq u < v) dv. \end{aligned}$$

From that, we deduce

$$\begin{aligned} \widehat{m_{ij}(T)} &= \ell_{ij}(\varphi) \int_0^T \frac{\mathbb{P}_\varphi(J(v) = i, N(u), S(u), 0 \leq u < v)}{\mathbb{P}_\varphi(N(u), S(u), 0 \leq u \leq T)} \\ & \quad \times \mathbb{P}_\varphi(N(u), S(u), v \leq u \leq T \mid J(v) = j) dv \end{aligned}$$

which completes the proof. ■